DOCUMENT RESUME

ED 164 623 95 TM 008 401

AUTHOR Tittle, Carol Kehr
TITLE Sex Bias in Testing: A Review with Policy Recommendations.
INSTITUTION ERIC Clearinghouse on Tests, Measurement, and Evaluation, Princeton, N.J.; Women's Educational Equity Communications Network, San Francisco, Calif.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Women's Program Staff.
PUB DATE 78
CONTRACT 300-77-0535
NOTE 58p.
AVAILABLE FROM Women's Educational Equity Communications Network, Far West Laboratory for Educational Research and Development, 1855 Folsom Street, San Francisco, California 94103 ($3.50)

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS Achievement Tests; Aptitude Tests; *Educational Testing; Females; Group Norms; *Guidelines; Interest Tests; Males; Occupational Tests; Policy Formation; Sex Differences; *Sex Discrimination; *State of the Art Reviews; *Test Bias; Test Reviews; Test Validity
IDENTIFIERS Information Analysis Products

ABSTRACT

Educational achievement tests, career interest inventories, and aptitude tests are reviewed for examples of sex bias, and changes in policy concerning the use of these tests are suggested. These suggestions are within the authority and responsibility of local and state educational administrators, teachers, counselors, parents, and students. The author concludes that guidelines concerning the review, selection, use, and interpretation of tests are especially needed for policy makers, since educational tests are neither inherently sex fair nor sex biased. Sources of guidelines for evaluating the fairness and the use of a test are discussed, as well as general professional standards for test construction. A number of examples from actual test items illustrate cases of bias against females--in achievement tests, male characters are mentioned more often and in more active roles; career interest inventories have separate male and female scales and norms, resulting in disproportionate counseling about career options; and aptitude tests may be written and interpreted according to sex role stereotypes. More stringent guidelines for test selection are strongly recommended whenever selection tests result in adverse impact. A bibliography is appended. (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED164623

WEECN
WOMEN'S EDUCATIONAL EQUITY COMMUNICATIONS NETWORK

Sponsored by the U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE/Office of Education
Under the Auspices of the Women's Educational Equity Act
Operated by FAR WEST LABORATORY FOR EDUCATIONAL RESEARCH AND DEVELOPMENT
1855 Folsom Street, San Francisco, California 94103
(415) 565-3032

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SEX BIAS IN TESTING: A REVIEW WITH POLICY RECOMMENDATIONS

Carol Kehr Tittle
University of North Carolina at Greensboro

1978

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Joseph A. Califano, Jr., Secretary
Mary E. Berry, Assistant Secretary for Education

Office of Education
Ernest L. Boyer, Commissioner
Women's Program Staff
Joan E. Duval, Director

National Institute of Education
Patricia Graham, Director
Central ERIC
Robert E. Chesley, Head

TM 008 401

ERIC ERIC CLEARINGHOUSE ON TESTS, MEASUREMENT, & EVALUATION
EDUCATIONAL TESTING SERVICE, PRINCETON, NEW JERSEY 08540

EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
CLEARINGHOUSE ON
TESTS, MEASUREMENT, AND EVALUATION (TM)

The material in this publication was prepared pursuant to a contract with the National Institute of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Prior to publication, the manuscript was submitted to qualified professionals for critical review and determination of professional competence. This publication has met such standards. Points of view or opinions, however, do not necessarily represent the official view or opinions of either these reviewers or the National Institute of Education.

WOMEN'S EDUCATIONAL EQUITY
COMMUNICATIONS NETWORK (WEECN)

This project has been funded with Federal funds from the Department of Health, Education, and Welfare, United States Office of Education, Women's Educational Equity Act Program, under contract number 300-77-0535. The contents of this publication do not necessarily reflect the views or policies of the United States Office of Education, the Department of Health, Education, and Welfare; nor does mention of trade names, commercial products or organizations imply endorsement by the United States Government. Material may be reproduced without permission, but credit would be appreciated.

• • •

DISCRIMINATION PROHIBITED—No person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance, or be so treated on the basis of sex under most education programs or activities receiving Federal assistance.

CONTENTS

Preface v

Introduction

- Why Examine Sex Bias in Educational Testing? 1
- Federal Laws and Regulations 2
- Needed: Procedures to Evaluate Equity in Testing 3
- Why is Sex Bias in Testing Important to Policy Makers in Education? 4

Sex Bias in Achievement Tests

- Overview: Content and Construct Validity 9
- Issue: Overt Sex Bias and Content Validity 10
  - Policy 13
- Issue: Item Bias and Construct Validity 19
  - Policy 23
- Issue: Test Bias and Construct Validity 23
  - Policy 24
- Policy Summary: Educational Achievement Tests 25

Sex Bias and Career-Interest Measurement

- Overview: Objectivity and Subjectivity in Measurement . 27
  - Types of Interest Inventories 27
  - The NIE Guidelines 28
- Issue: Face Validity 29
  - Policy 29
- Issue: Same Scales for Females and Males 30
  - Policy 31
- Issue: Career Options Suggested to Females and Males .. 32
  - Policy 32
- Issue and Policy: Minorities and Mature Women and Interest Inventories 33
- Issue: The Purpose of Interest Inventories and Validity .. 33
  - Policy 34
- Issue: Counselor Use of Interest Inventories 35
  - Policy 36

Aptitude Measures and Sex Bias

- Overview 39
- Issue: Overt Sex Bias 40
  - Policy 40
- Issue: Sex Differences in Performance and Their Interpretation 41
  - Policy 42
- Issue: Adverse Impact of Selection Tests 43
  - Policy 45

Concluding Remarks 47

References 49

PREFACE

Sex discrimination has been prohibited by law in educational institutions that receive federal support. Yet, there is still concern that many educational policies, procedures, and practices reinforce sex-role stereotypes and reflect prejudiced views of women's achievements, abilities, and interests. Educational tests have been examined for bias against women by a variety of procedures (Tittle, 109), as tests in general have been examined for bias against minority groups (Flaugher, 33). The variety of procedures used in studies indicates that there is not one procedure or definition that will quickly tell the policy maker, test user, or test publisher that an educational test is or is not biased against women. Rather, there are currently being developed a series of guidelines encompassing procedures that will need to be considered to permit the statement that a test and its use are sex-fair. Or, said in another way, that the test and the test use are as free from sex bias as it is possible to determine, given the present state of the art in identifying aspects of discrimination and in the field of testing. This monograph reflects the state of the art in both testing and analyses related to educational equity. Both provide sufficient guidance to inform policy, improve educational practice, and provide sex-fair instruments and test use.

Carol Kehr Tittle is currently a professor in the School of Education, University of North Carolina at Greensboro. She was formerly an associate professor at the Center for Advanced Study, Graduate School and University Center, City University of New York, New York City. She is author of Women in Testing and is an editor for the forthcoming book The Handbook of Vocational Education Evaluation.

INTRODUCTION

WHY EXAMINE SEX BIAS IN EDUCATIONAL TESTING?

Educational testing is a common experience in the school life of American children. Testing is carried out at all levels of the educational system. From a concern at the national level for the status of basic knowledge of American students and adults, as monitored by the National Assessment for Educational Progress, to the state-level testing of minimum competencies for graduation or assessment of statewide progress in education, to the use of tests in local school systems for assessing and diagnosing student progress, most school children and the majority of adults today will have had the experience of taking an educational test. Tests are also widely used in evaluations of federal programs that allocate funds to the states for various educational programs.

Other major areas of test use are in the selection of students for college admissions and in career guidance and counseling. Tests are also used in the selection of students for occupational programs. School systems and states also use tests to certify teachers and to select employees. An estimate of the types of tests administered annually was made by Holman and Docter (51). They listed three main areas of testing programs and their proportion of total test use: 1) educational achievement testing (65 percent); 2) testing for selection and placement (30 percent); and 3) testing in counseling, guidance, and clinical work (5 percent of tests used). Some indication of the volume of tests administered is found in a recent estimate of 3.5 million interest inventories scored annually by major test scoring services (111). Holmen and Docter estimated that 200 million achievement test forms and answer sheets were used annually in the United States as of 1972.

In higher education, where tests are used in admission procedures, other forms of educational testing will also become more prominent. The City University of New York, for example, has recently instituted a policy of assessing student competence in writing and reading at the transition point between the sophomore and junior years. This policy will affect about 150,000 students. Educational testing then, from the kindergarten and first-grade level of Title I testing, through testing for minimum competency standards for high school graduation, to college entrance and minimum competencies for transition points in college, to the use of interest measures to assist career choice, is widespread throughout the American educational system.

The increasing numbers of tests being developed and the large numbers of students tested, half of whom will be women, has made educational testing a subject of examination by those concerned with equality of opportunity for women and educational equity. Since the test content becomes a part of the school's materials, just as textbooks and beginning readers are, they help to form the view that students have of themselves. In particular, they help to reinforce and illustrate the views our culture holds of appropriate roles for men and women. These roles for women and men are conveyed in a number of spheres: in the home, in school, in male-female interactions, in child rearing, and in occupational settings.

In addition to the fact that tests are part of the educational setting for students, they are important in another way. This other evidence of their importance is found in the effects of tests on students, parents, and teachers. The Russell Sage Foundation has funded a series of studies that have examined attitudes held about intelligence tests and teachers' views of their preparation for understanding tests. In a survey about American beliefs and attitudes about intelligence, Brim et al. (7) found that almost 80 percent of public school students believe that intelligence tests are somewhat or very accurate. There is some evidence that these views are held generally and extend to the areas of college entrance

testing and the assessment of school achievement (61).

Goslin (35) conducted a survey examining teacher attitudes toward standardized tests. He found that they tend to view standardized tests as relatively accurate measures of a student's intellectual potential and achievement; that teachers see the kinds of abilities measured by standardized tests as important determinants of the subsequent academic success of children; and that they believe that considerable weight should be given to test scores, along with school grades, in making decisions about special classes, college admissions, and so on. There is an indication of an internal consistency in the belief systems of some teachers concerning tests and their use. Teachers who expressed confidence in the accuracy of standardized tests also felt that they measured the qualities necessary for success; they also believed that the abilities measured were, to a significant degree, innate, rather than learned. These teachers tended to have had more contact with tests and more formal training in psychometrics.

FEDERAL LAWS AND REGULATIONS

In addition to the widespread use of tests, their place in the context of education for students, and the likely beliefs of their value held by teachers, there are other current reasons for examining educational tests for educational equity for women and educational policy. Federal law now provides regulatory and legal pressures for equity and fairness in testing. Three examples are Title IX, the Uniform Guidelines on Employee Selection Procedures, and the Vocational Education Act of the Education Amendments of 1976. Title IX, in the Education Amendments of 1972, has the regulations that most widely affect all levels of educational practice. Title IX prohibits discrimination on the basis of sex against most adults employed in educational settings and most students. The same benefits and opportunities for job advancement are to be offered men and women, and boys and girls are to receive the same instruction and treatment without regard to their gender. Although curriculum materials are excluded from Title IX, the use of tests and counseling are not. Tests and materials used by counselors and teachers in guidance must be nondiscriminatory. If this is broadly interpreted, achievement, aptitude, and interest tests all fall within the purview of Title IX.

Schiffer (95) has described the legal regulations that are related to selecting and using interest inventories under Title IX and the 14th Amendment to the Constitution (the Equal Protection Clause). Title IX of the Education Amendments Act of 1972 has specific requirements for eliminating bias in test use and counseling for every school that receives federal funds. Title IX provides:

> No person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity receiving federal financial assistance....

One section of the regulation is particularly relevant to schools in using interest inventories and counseling for career selection. This is 45 C.F.R. 586.36: Counseling and Use of Appraisal and Counseling Materials.

> (a.) Counseling. A recipient shall not discriminate against any person on the basis of sex in counseling or guidance of students or applicants for admission. (b.) Use of appraisal and counseling materials. A recipient which uses testing or other materials for appraising or counseling students shall not use different materials for students on the basis of their sex or use materials which permit or require different treatment of students on such basis unless such different materials cover the same occupations and interest areas and the use of such different materials is shown to be essential to eliminate sex bias. Recipients shall develop and use internal procedures for ensuring that such materials do not discrimate on the basis of sex. Where the use of counseling, test or other instruments results in a substantially disproportionate number of members of one sex in any particular course of study or classification, the recipient shall take such action as is necessary to assure itself that such disproportion is not the result of discrimination in the instrument or the applications.

Similar ideas appear in the Title IX regulations regarding discrimation on the basis of sex in admission and recruitment. The specific prohibitions in this section of Title IX include the following:

> (1) In determining whether a person satisfies any policy or criterion for admission, or in making any offer of admission, a recipient to which this subpart applies shall not: (i.) give preference to one person over another on the basis of sex, by ranking applicants separately on such basis, or otherwise; (ii.) apply numerical limitations upon the number or proportion of persons of either sex who may be admitted; or (iii.) otherwise treat one individual differently from another on the basis of sex.
>
> (2) A recipient shall not administer or operate any test or other criterion for admission which has a disproportionately adverse effect on persons on the basis of sex unless the use of such test or criterion is shown to predict validly success in the education program or activity in question and alternative tests or criteria which do not have such a disproportionately adverse effect are shown to be unavailable. (Federal Register, 1975, p. 24140)

These regulations are similar to the regulations in the Uniform Guidelines on Employee Selection Procedures (Federal Register, December 20, 1977). The Uniform Guidelines represent a consensus among four federal agencies: the Civil Service Commissions, the Equal Employment Opportunity Commission, the Department of Justice, and the Department of Labor. In the Uniform Guidelines, the definition of discrimination that is used with regard to employment decisions is also on the basis of adverse impact. Employment decisions are broadly defined; they include, but are not limited to, hiring, promotion, demotion, membership (for example, in a labor organization), referral, retention, licensing, and certification. Other selection decisions, such as selection for training, may also be considered employment decisions. Discrimination is defined as the use of any selection procedure that has an adverse impact on the hiring, promotion, or any other employment or membership opportunities of members of any racial, ethnic, or sex group unless the procedure has been validated in accordance with the Guidelines. Adverse impact is defined as a selection rate for any racial, ethnic, or sex group that is less than four-fifths (80 percent) of the rate for the group with the highest rate. This selection rate will be generally regarded by the federal enforcement agencies as evidence of adverse impact. Smaller differences in selection rate may constitute adverse impact, where they are significant in both statistical and practical terms. It is interesting to note that selection procedure is very broadly defined. Selection procedures are defined as including the full range of assessment techniques from traditional paper and pencil tests, performance tests, training programs or probationary periods, and physical, educational, and work experience requirements, to informal or casual interviews and unscored application forms.

The Education Amendments of 1976, in the Vocational Education Act (VEA), also sets forth the policy of equal access for all minorities and women to programs under the legislation and requires states to describe the specific actions taken to overcome sex discrimination. States are also to specify the incentives adopted to encourage enrollment of both women and men in nontraditional courses of study. State plans are to include model programs developed to reduce sex bias and sex stereotyping in training programs and placement in all occupations (Section 104.187, Federal Register, Monday, October 3, 1977). Federally funded programs, as well as the programs of local districts, must be concerned with eliminating discrimination based on sex and with providing sex-fair counseling and guidance activities and materials, including educational tests.

NEEDED: PROCEDURES TO EVALUATE EQUITY IN TESTING

The various regulations concerned with sex bias and discrimination provide general principles. A few regulations are specific to sex bias in educational testing, e.g., the use of career-interest inventories in counseling and employment testing. How-

ever, there is not presently consensus on the full set of procedures that would define a sex-fair (or unbiased) test and its use in all educational settings. The major sections of this publication provide guidance to policy makers and test users based on the principles in the various regulations.

The remainder of this publication has two goals. The first is to present illustrative statistics indicating why administrators and policy makers should be concerned with educational tests. The why is the evidence underlying the need to challenge discrimination and the limiting of options for women. Secondly, and the main focus of the presentation, is to raise a series of critical issues for each of the major types of educational tests--achievement tests, career-interest inventories, and aptitude measures. The purpose of describing the issues is to indicate needed policy, procedures that can formalize policy, and the why of the procedures. These issues will help to define and suggest at least some of the procedures that are necessary to make a judgment of the sex-fairness of a test of educational achievement, career interest, or aptitude used in selection for specific vocational courses, college admissions, or employment.

The procedures that would ensure sex-fairness for different groups of women are not well-identified. Issues in the testing of minority women and older women are not treated separately in the present work. As a reviewer of an early draft commented, the issues of racial bias are at least as complex as those considered here, and the combined bias that faces minority women has nowhere been adequately considered. This is as true in the sex bias literature as in race/ethnic bias literature. The reader may find additional concerns identified by consulting a resource such as the Psychological Testing of American Minorities (94).

WHY IS SEX BIAS IN TESTING IMPORTANT TO POLICY MAKERS IN EDUCATION?

Earlier it was suggested that it is important to look at educational tests because they are part of the student's context in education and because they are widely used. Underlying the present concern with educational tests is a thesis on the relationship between education for women, the degrees they obtain, and their occupational entry and career paths, a thesis that makes tests a concern to policy makers in education.

There is a riddle that is sometimes posed to illustrate the strengths of our stereotypes about women. The riddle begins with a father and son driving on their way to the next town. They have a car accident and the father is killed instantly. The son is very seriously injured and is taken to a hospital. The son is immediately taken up to the surgical ward. The surgeon comes in and says, "I cannot operate on this boy, he's my son." How do you explain the riddle? Listeners often puzzle over how the father could be dead and still be there to operate on his son! The source of the puzzlement is understandable. There are few women surgeons in this country, and television and other media do not show women as surgeons and rarely as doctors. While listeners may be aware that more women are entering medical schools today, they still may not immediately find the logical solution to the puzzle.

The reactions to the riddle are no accident, but derive from a historical view of women and can be illustrated by historical views of the nature of women and men, which present a picture of what women are thought to be. The following quotations are some of these views of women and men,[1] that limit our perceptions of individuals.

> We may thus conclude that it is a general law that there should be naturally ruling elements and elements naturally ruled...the rule of the freeman over the slave is one kind of rule; that of the male over the female another...the slave is entirely without the faculty of deliberation; the female indeed possesses it, but in a form which remains inconclusive... (Aristotle)

> The chief distinction in the intellectual powers of the two sexes is

---

[1] The quotations that follow are taken from Barbara Deckard, The Woman's Movement: Political, Socioeconomic and Psychological Issues. New York: Harper and Row, 1975, Pp. 1-4.

shown by man attaining to a higher eminence, in whatever he takes up, than woman can attain--whether requiring deep thought, reason, or imagination or merely the use of the senses and hands. (Darwin)

As much as women want to be good scientists and engineers, they want, first and foremost, to be womanly companions of men and to be mothers. (Bruno Bettelheim)

Nature intended women to be our slaves;...they are our property, we are not theirs. They belong to us, just as a tree that bears fruit belongs to a gardener. What a mad idea to demand equality for women! ...Women are nothing but machines for producing children. (Napoleon Bonaparte)

These quotations are reflected in the history of education for women in the United States. As Mathews (70) summarized this history, it was not possible for women to attend college in this country for 200 years after the founding of Harvard in 1636. Oberlin became the first college in the nation to admit women; in 1855 Elmira Female College was founded; and in rapid succession after the Civil War the others later known as the Seven Sisters were founded. The first state-supported college for women in the world was chartered in 1884 and is now known as the Mississippi State College for Women. After the Civil War, the number of women that were accepted into previously all-male institutions seems to be inversely related to the economic strength of the college. Established colleges in the East that had heavy endowments retained their exclusively male student body. Colleges in the West and denomination colleges and state universities exhibited less resistance to women as students. Between 1870 and 1890 the number of colleges that admitted women almost doubled, and the number of female college graduates increased five-fold. By 1900 graduate and professional schools were opened to highly motivated women, for the most part. After 1920, the number of women college graduates increased and continued to rise to the approximately 40 percent of the graduating classes today.

TABLE 1

Distribution of Females and Males in Vocational Education for Each Program Area, 1972

	FEMALES			MALES		
	Gainful (Excl. Homemaking) (%)	*Incl. Homemaking* (%)	*N*	*Gainful* (%)	*Incl. Homemaking* (%)	*N*
Agriculture	1	1	48,153	17	17	848,307
Distributive education	8	5	290,020	7	7	350,403
Health	8	4	285,071	1	1	51,581
Home economics: gainful	7	4	240,948	0.1	0.1	39,018
Office	51	28	1,796,387	11	11	555,491
Technical education	1	1	33,006	6	6	304,063
Trades/industry	8	4	279,680	43	43	2,118,288
Special program*	17	9	582,715	15	15	721,904
Total: gainful only	101	...	3,505,128	100	...	4,931,284
Home economics: homemaking	...	45	2,916,987	...	0.5	248,745
Total gainful and homemaking		101	6,422,115		101	5,180,029

SOURCE.—Calculated from Bureau of Adult, Vocational and Technical Education, *Summary Data: Vocational Education, Fiscal Year 1972* (Washington, D.C.: U.S. Department of Health, Education and Welfare, Office of Education, May 1973), p. 2.

*Includes prevocational, prepostsecondary, and remedial programs.

Roby (92) has described the recent increases in enrollments in vocational education. As college enrollments have leveled off, state vocational and technical education enrollments increased from roughly 7.5 to 11.6 million from 1968 to 1972. Adding to the influence of vocational education was the Higher Education Act of 1972 authorization of $950 million for post-secondary occupational (i.e., vocational) education. This gave an impetus to institutions of higher education to further expand vocational education.

However, the statistics for women in vocational education are less than encouraging (92). As of 1972, there were 6.4 million women and girls enrolled in public vocational programs across the country. Of these girls and women, 45 percent were being trained in homemaking and home economics and another 28 percent in office practices. Project Baseline Report data (63) indicates that there have been changes in the various areas of vocational education. However, the data continues to show that very few women are being trained for the 20.1 million jobs that the National Planning Association estimates will occur in what have been viewed primarily as male occupations, including the better-paying trades, industrial, and technical jobs for which high schools now offer vocational courses with entry-level skills preparation.

TABLE 2

Enrollment in Vocational Education Programs FY1975

PROGRAM	*Total Enrollment*	*No. of Males*	*Males as % of all students enrolled*	*No. of Females*	*Females as % of all students enrolled*
Agriculture	814,427	737,573	90.56	76,854	9.44
Dist. Educ.	624,835	331,598	52.98	294,237	47.02
Health	461,512	100,250	21.72	361,262	78.28
Consumer & Homemaking	2,766,482	433,383	15.67	2,333,099	84.33
Home Economics	286,612	48,435	16.09	238,177	83.10
Office	1,940,822	513,874	26.58	1,426,948	73.52
Technical, Trade and Industry	2,107,741	1,824,748	86.57	282,993	13.43
Special Programs	1,858,050	1,444,324	61.59	713,726	38.41

SOURCE.—See Lee and Fitzgerald, 1976.

Similar sex-segregated distributions occur in both higher education and labor force participation in general. Women work in all occupational categories, but they are concentrated in fewer occupational categories than men (113). Women constitute 53 percent of the civilian non-institutional population 16 years old and over, but more women are employed part time (64 percent), and women constitute only a third (31 percent) of all persons employed in professional-technical and non-farm managerial administrative occupations. Women are predominant (68 percent) in the persons employed in clerical-sales occupations. The greatest rate of increase of women in the labor force between 1960 and 1975 is for women with young children. The increases were from 15 to 33 percent among women with children under the age of 3, and from 25 to 42 percent among women with children between the ages of 3 and 5 (1). Equity in education and earnings are important to all working women. Although women make up 42 percent of the work force, they receive only 25 percent of the total earned wages going to American workers (82). The average woman earns 60 percent of what the average man earns--a smaller share than 20 years ago, when woman's paycheck averaged 63 percent of the average man's (80).

In higher education the proportion of women in science is small and the proportion drops at each higher level of degree, salary, academic rank, and administrative responsibility (Vetter, 115). Among almost 207,500 science and engineering Ph.D.s in the U.S. labor force, 92 percent are male. The proportion of women enrolled and graduating in these fields was higher in the 1920s than in any decade since, but now appears to be rising. In the field of chemistry for example, according to Vetter, women have earned 19 percent of the bachelor's degrees, 20.8 percent of the masters, and 7.3 percent of the doctorates since 1960. Women earned 11.17 percent of the chemistry doctorates in the period 1973-1976 (National Research Council, Women and Minority Ph.D.s in the 1970s). However, at institutions awarding the doctorate in 1973, only 2 percent of the chemistry faculty above the level of instructor were women, and only 14.8 percent of federally employed chemists at all degree levels were women.

There is a lack of data showing significant differences in talent that would account for the discrepancy in education, vocational entry, and career attainment. The most comprehensive review of sex differences in the psychological literature has been carried out by Maccoby and Jacklin (67). At the conclusion of their detailed review of the psychological literature on sex differences, they could find few consistent differences in performance. Even in the few differences found, they were inclined to emphasize considerable overlap of the distribution of male and female abilities or talents on whatever psychological dimension was being measured. There are, for example, many girls with high-level skills in mathematics, just as there are many boys with high-level skills in the verbal area, two traditionally stereotyped areas of men's and women's skills. There are more than enough women to fill engineering schools if women's talents were developed through the requisite early training and interests. According to Maccoby and Jacklin:

> Women are now considerably underrepresented in engineering in terms of any criterion by which potential talent can be measured. We have no wish to push women toward careers that do not attract them. At the same time we belive it would be a grievious injustice to establish formal or informal quotas that would exclude any women with the requisite talents and interests. We are discussing quotas that exclude women because, historically, women have been excluded from training for high status careers more frequently than men, but of course the argument applies in both directions.

Applicants or students should be assessed and counseled on the basis of their measurable talents, not on the basis of probabilities of attainment on sex.

There appear to be some documented sex differences in performance, including the visual-spatial skills areas, mathematics aptitude and achievement at later stages of education, verbal performance, and perhaps some social behaviors such as aggressiveness. However, in the research surveyed, there are problems in establishing comparability of groups in experience and education levels, of possible bias in

measures, and of measuring particular abilities and social skills in a wide variety of contexts. There is a great danger in stereotyping males or females on the basis of mean differences in group performance in view of extensive overlapping among the groups on distributions in all abilities measured.

One area in which it appears there are differences in the variables measured is in career interests. There are differences among female and male responses to existing item pools. However, the discussion in the section on interest measurement points out that these data are limited by being based on sets of items that have shown sex differences over long periods of time and currently may not be related to the measurement of interests of specific occupational groups.

Policy makers need to be concerned that tests may function to reinforce existing performance patterns. For example, if interest measures do not provide the same sets of career options to males and females, then they are likely to reinforce existing occupational distributions in the labor force. Similarly, the results of aptitude measures that students take with differential experience prior to the test may also serve to reinforce existing stereotypes of skills of females and males. The results of tests of educational achievement in mathematics may be misleading for women and those who are using the test scores, if the amount of math experience that females and males have is not equivalent. These examples are examined in more detail in the issues sections for each of the major types of educational tests.

SEX BIAS IN ACHIEVEMENT TESTS

OVERVIEW: CONTENT AND CONSTRUCT VALIDITY

The simplest definition of validity is captured in the question: Are we measuring what we think we are measuring? If teachers construct a test to assess the achievement of their ninth grade students in algebra, or if they want to select a standardized test to measure this achievement, they are concerned with content validity. Typically, content validity is assessed by examining the match between the curriculum and the test content. For standardized achievement tests, test publishers develop a set of specifications that define the domain of content to be measured by the test. Typically, this content domain specifies both the substance and the process (or task behavior) to appear in the test. By substance, in arithmetic achievement for example, the test specifications might be concerned with whether students are answering questions on fractions or whole numbers. The process dimension refers to whether they are asked to add, subtract, multiply, or perform other types of manipulations on the substance or content. Content validity, then, is concerned with the representativeness or sampling adequacy of the test items for the domain to be measured and is typically assessed by judgment. The persons selecting a test will provide either formal or informal sets of judgments as to the content validity of the test for their particular group of students and their curricula.

Strictly speaking, if a student does not answer questions on a test, the resulting low score is a valid report in one sense. The danger in drawing this conclusion immediately was described by Messick (75). Suppose a deaf pupil has been given a spelling test by dictation. Although the low score is a valid report that the pupil did not spell from dictation, the inference that the pupil lacked the ability to spell those words is much more tenuous. One can infer from correct performance that the student possessed the requisite abilities. However, to make the inference of inability or incompetence from the absence of correct performance requires the elimination of a number of plausible rival hypotheses.

Psychologists have used the term construct validity to make more apparent the nature of inferences that are drawn from test scores. If a student has a low score on a test, we want to know the meaning of the test score. Does the absence of correct performance mean a student has not had an opportunity to learn the material? The construct validity of achievement tests is concerned with demonstrating that the opportunity to learn variable is the main reason that students may not have achieved a desired level. Plausible rival hypotheses for lack of a correct performance need to be ruled out. Status characteristics such as sex, ethnicity, and socioeconomic status should not be the major variables accounting for individual differences in student performance on achievement tests.

Another sense of bias is the presentation of test material that is sex-role stereotyped and not a fair representation of women's status and achievements in our culture and in history. Sex-fairness in representation of women in occupational roles and scientific achievements is the first issue discussed below and is one aspect of the "face validity" of an achievement test. The effect of sex-fairness of representation of women may or may not be reflected in achievement test performance. (Evidence bearing on this issue is described later.) Whether or not "biased" content is reflected in performance is irrelevant to the larger issue of fairness of representation. Test content, to the extent that illustrative materials can represent both sexes fairly, should be representative and face valid on a basic principal of justice or fairness.

The efforts to examine item or test bias from a statistical or empirical viewpoint have an important role to play and are another part of the procedures to eliminate sex bias and provide a sex-fair test. In the long run, however, because there are still many culturally different experiences for boys and girls, it will be some time before all differences between the sexes in performance on achievement tests are eliminated. However, the item bias studies can help to reduce the amount of irrelevant differences. Empirical study of items can reduce at least part of the sex differences in performance that are not attributable to direct instructional experience. Some differences may be attributable to earlier experiences. For example, boys are more likely to have experience with mechanics, electrical repair, and so on. Items that embed mathematical processes in these contexts may be easier for boys than for girls because of their greater familiarity with the type of situation in the item. If the same mathematical process can be measured with a sex-neutral content, one on which boys and girls on the average perform similarly, this type of item is not "sex biased" and is preferable if it meets the test specifications for a sex-fair and content valid test.

The discussion below and many of the studies cited are based on standardized (norm-referenced) tests. This is because many of the studies to examine sex or minority group bias are conducted in large-scale assessments of pupil achievement that use standardized tests. However, the issues and policies below also apply to criterion or objectives-referenced achievement measures. The examination of items for their face validity, or fairness of representation, and for statistical or empirical bias is equally mandatory for both norm-referenced achievement tests and for objectives-referenced achievement tests.

ISSUE: OVERT SEX BIAS AND CONTENT VALIDITY

An earlier review (112) of the portrayal of women in educational tests included this summary:

> Women are portrayed almost exclusively as homemakers or in the pursuit of hobbies (e.g., "Mrs. Jones, the President of the Garden Club..."). Young girls carry out "female chores" (e.g., father helps Betty and Tom build a playhouse; when it is completed, "Betty sets out dishes on the table, while Tom carries in the chairs...").
>
> In numerous activity-centered items, boys are shown playing, climbing, camping, hiking, taking on roles of responsibility and leadership. Girls help with the cooking, buy ribbon and vegetables, and when participating in any active pursuit, take the back seat to the stronger, more qualified boys (e.g., Buddy says to Clara, "Oh, I guess it's all right for us boys to help girls. I've done some good turns for girls myself, because I'm a scout.").

Items in achievement tests have conveyed the impression that the majority of professions are closed to women. A reading passage about the presidency of the United States discussed the president's characteristics and qualifications and included the statement, "In the United States, voters do not directly choose the man they wish to be president." Routinely, teachers are described as females, while professors, doctors, lawyers, and presidents of companies are listed as male. These statements apply to the achievement tests as reviewed for overt bias in 1974. The tests were not unknown tests, but were the achievement tests published by the major test publishing companies. And the tests examined carried publication dates from 1964 to 1971. The analysis of the achievement tests showed the same results as the analysis of other types of educational materials that have been examined for discrimination against women. (See, for example, 59, 34, 93, 53, and 37.)

These studies of instructional materials, as well as the study of educational achievement tests, provided extensive documentation on one type of overt sexism, that of sex-role stereotyping in content. A second way of examining overt sexism is to look at language usage. Another estimate of content bias may be obtained by determining if males are referred to more often than females. Reading comprehension, social studies,

science, and other tests may well cite male novelists more frequently than female, describe male scientists' activities more frequently, and so build, or rather reinforce, the view of the occupational world as male oriented.

There is some evidence that sex bias in content may largely arise through content selection. The preparation of the new American Heritage Dictionary involved the computer analysis of 10,000 500-word samples from 1,000 of the most frequently used instructional publications in representative schools across the country (68). The analysis of these school materials showed evidence of male orientation. The word boy or boys appears 4,700 times versus 2,200 for girl or girls. Of the 20 given names most frequently used, 13 were male and 7 were female. These are specific instances of male orientation in school materials. Several writers have discussed the general male orientation of the English language, in what appears to be sex-typed use of language (103, 59, 108).

An estimate of the weighting of content toward males or females can be obtained by computing rations of frequency of usage of male nouns and pronouns to female nouns and pronouns. One factor that apparently contributes to bias is that the English language has no singular pronouns equivalent to the plurals they, their, and them. Common usage attributes "maleness" to most occupations, for example, the carpenter..."he," the counselor..."he," and the writer (author) ..."he."

A study by Tittle, McCarthy, and Steckler (112) examined whether overt bias arose from content selection or whether it could be attributed to the common use of language and the generic nouns and their pronoun referents (i.e., references to such nouns as mankind, chairman, fisherman). In order to separate the content selection and usage factors, two ratios were computed: 1) the ratio of the count of the frequencies of male nouns and pronouns to female nouns and pronouns, using only regular nouns and pronouns; and 2) the ratios of counts of all male nouns and pronouns to all female nouns and pronouns including generic nouns and pronouns, a count labeled all nouns and pronouns.

These ratios provide the basis for one examination of sex bias in language usage in achievement tests. Eight of the most frequently administered achievement test batteries were analyzed and the major findings were:

(1) There were few differences between the conclusions drawn based on the ratios using all nouns and pronouns and those using regular nouns and pronouns only (i.e., excluding generic noun and pronoun referents). Content bias in favor of males did not appear to be primarily a function of word choice, but rather of content selection.

(2) With one exception, each test battery showed a higher frequency of usage of male nouns and pronouns than of female nouns and pronouns. The range of the ratios was from .86 (slightly more female references than male) to a high of 14 (14 times more use of male nouns and pronouns than female).

(3) The number of subtests in each achievement test series with ratios below 1.00, where more female than male nouns and pronouns were used, ranged from none (out of 7 subtests) to 5 subtests (out of 9). The majority of achievement test batteries show few subtests with ratios at or below 1.00, another indication of the imbalance in the references to females and males in the achievement tests analyzed.

(4) The extent of the imbalance was shown in the five subtests with the highest ratios of frequencies of male nouns and pronouns to female--84:1; 84:1; 69:1; 41:1; and 33:1. These figures indicate that in two subtests 84 male nouns and pronouns were counted and only one female noun or pronoun.

These findings were also confirmed in a study carried out by Donlon, Ekstrom, Lockheed, and Harris (22). Thus the use

of language and sex-role stereotypes confirms the under-representation of women generally and an over-representation of women in traditional settings. It is possible to improve the counts and findings as shown in the development of the 1978 edition of the Metropolitan Achievement Tests (MAT). Jensen and Beck (54) reported on a gender balance analysis of the new MAT. The publisher balanced the presentation of females and males and portrayed both sexes in less stereotypic roles. There was a marked difference from the 1970 edition to the 1978 edition of the MAT. Although some tests retained an imbalance, the overall balance was improved, changing from a median ratio of male nouns and pronouns to female nouns and pronouns of 2.95, to a ration of 1.1. The study of the MAT also examined traditional and nontraditional views of women in four categories: occupations, activities, roles, and emotions. One interesting finding of this analysis was that there was a greater tendency for females to be better represented in both traditional and nontraditional categories occupations and activities, than for males.

Of some interest in relation to "sex bias" in language in achievement tests is the relationship between sex bias in language and performance on an achievement test. As noted earlier, there are good reasons to change the overt sexism in language in achievement tests as the MAT has done, regardless of whether or not there is a relationship to test performance. On the other hand, the topic is of some interest and there have been studies of these relationships. Plake, Hoover, and Loyd (84) looked at the differential item performance by sex on the Iowa Tests of Basic Skills (ITBS) for grades 3 through 8. Although some items were found on which there were different performances by males and females (statistically significant), the differences were not practically significant. Plake indicated that her results provided little support for the idea that test content as it relates to sex-role stereotyping or the frequency of sex-identified nouns and pronouns affected performance on mathematics or other test items.

Donlon et al. (22) did find some evidence of differential performance by females and males for items where the test content reflected more male nouns and pronouns than female nouns and pronouns. The measure that the Donlon et al. study used was sex difference in passing the item (male percent passing minus female percent passing) correlated with total male references and total female references. This measure showed a few differences except for the STEP reading test at grade 10.[2] Here there were moderate correlations between sex differences in passing and total male references and total female references. More generally stated, there was a very moderate but significant tendency for females to do better on items that contained more female references. Other findings of interest in the Donlon study were that females and males were highly similar on such factors as their rate of work in completing the test and the number of items omitted.

In addition to the studies by Plake and Donlon et al., three other studies have been reported that examine sex differences in performance on tests of mathematics as a function of the item context. Two studies used pupils at the elementary school level and looked at sex differences in performance for items in which the activities were familiar or stereotypically for men or women. King and Blount (60) used a teacher-developed test and reported differences in performance of sixth grade students in the direction the opposite of prediction. That is, the girls did better on items with a masculine orientation. Mowder and Chase (78) reported no sex difference in performance of third grade pupils when responding to sex-typed items. Sixth grade data showed an interaction effect, with girls doing better than boys on the "male-bias" test. McCarthy (71), however, found more systematic differences in performance when using a sample of high school students in grades 10 to 12.

---

[2] Items were analyzed for the following tests: California Achievement Test (Level 5, Form A), Iowa Tests of Basic Skills (Levels 11 and 14, Form 6), Metropolitan Achievement Test (Grade 12, Form F), and The Sequential Tests of Educational Progress (Grade 10, Series 11).

Her results are more in line with the general trend in the research literature toward no differences in math performance between the sexes in the early levels and some findings of sex differences at the later, high school grade levels.

McCarthy constructed an item pool, varying the item context and holding the mathematical process the same. She had a subsample of students rate each item on a scale of 1 to 5, 1 being of great familiarity to males, 5 being of great familiarity to females. Items selected or the final test were definitively rated male, female, or neutral. She used the relationship between passing the item and performing well on the test to select the 26-item "best" tests for the Total group, for Males only, and for Females only. With this approach, the items selected for the Total group included 10 male items, 4 female items, and 12 neutral items. The items selected for the Male group included 11 male items, 6 female items, and 9 neutral items. The items selected for the Female group included 2 male items, 14 female items, and 10 neutral items. The item overlap was: the Total group and Male group had 20 of 26 items the same (77 percent overlap); the Total group and Female group had 15 of 26 items the same (58 percent overlap); and the Male group and Female group had 13 of 26 items the same (50 percent overlap). An analysis that controlled for attitude toward mathematics and verbal ability did not reduce the significance of the differences found: males scored higher than females on the Total group test and Male test; females scored significantly higher on the Female group test.

McCarthy's findings are consistent with earlier studies indicating that by the high school or college level, women's performance on tests of mathematics involving problem solving may be affected by their familiarity with (or attitude toward) the context in which the problem is posed. Milton (76) reported a series of five studies using high school and college students in which he showed that problem-solving differences between the sexes were reduced (but not eliminated) when the problems were written in content appropriate to the feminine role.

A related study was conducted with an "aptitude" test. Strassberg-Rosenberg and Donlon (104) examined sex differences in performance on mathematics items on the Scholastic Aptitude Test and found that more items were "biased" (had higher percent passing) on the average in favor of males on the regular math items and data sufficiency geometry items. Items biased in the female direction were algebra, regular math items, and five miscellaneous regular math items. These findings were similar to Donlon's (21) earlier study of SAT items, in which he found items with algebra content to be easier, on the average, for women than the other items. His earlier study also found a definite masculine tenor to the content of 17 items with real-world referents.

The summary of what is known about overt sex bias in achievement tests can be stated as follows: studies have documented the bias in favor of males in selection of item content and that there are sex-role stereotyped views of women and men. There is some evidence that familiarity with item context may affect performance in mathematics tests at the high school age and also that more frequent references to women may affect performance at the high school level. Both of these latter findings are based on few studies and need to be replicated. However, there are clear implications for policy from the findings that are well documented.

Policy

Test publishers are taking steps to eliminate overt sexism from tests and the accompanying test interpretation materials. However, the policy decisions at schools and other educational groups selecting achievement tests should focus on developing standards in several areas for use in test reviews. This section lists several of the types of categories and procedures that have been used by publishers or in studies of tests and instructional materials to examine sex-role stereotyping and sex bias in language usage. These can be adapted for school use.

A readily used form was developed by Carol Jacobs and Cynthia Eaton (53). The form was originally developed to evaluate sexism in readers for the elementary school and is illustrated in Figure 1.

FIGURE 1

Tally Form for Reviewing Educational Materials
(Jacobs and Eaton, 53)*

EVALUATING SEXISM IN READERS

	MALE	FEMALE
1. Number of stories where main character is:	_____	_____
2. Number of illustrations of:	_____	_____
3. Number of times children are shown:		
(a) in active play	_____	_____
(b) using initiative	_____	_____
(c) displaying independence	_____	_____
(d) solving problems	_____	_____
(e) earning money	_____	_____
(f) receiving recognition	_____	_____
(g) being inventive	_____	_____
(h) involved in sports	_____	_____
(i) fearful or helpless	_____	_____
(j) receiving help	_____	_____
4. Number of times adults are shown:		
(a) in different occupations	_____	_____
(b) playing with children	_____	_____
(c) taking children on outings	_____	_____
(d) teaching skills	_____	_____
(e) giving tenderness	_____	_____
(f) scolding children	_____	_____
(g) biographically	_____	_____

5. In addition, ask yourself these questions: Are boys allowed to show their emotions? Are girls rewarded for intelligence rather than for beauty? Are there any derogatory comments directed at girls in general? Is mother shown working outside the home? If so, in what kind of job? Are there any stories about one-parent families? Families without children? Are baby-sitters shown? Are minority and ethnic groups treated naturally?

*Reprinted by permission of Today's Education, The Journal of the National Education Association.

Table 3

Categories Used by Jacklin to Examine Sex Bias
(Saario, Jacklin, and Tittle, 93*)

CATEGORIES

1. Main and secondary characters

2. Type of environment:
   Home
   Outdoors
   Place of business
   School

3. Behavior exhibited:
   Nurturant (helping, praising, serving)
   Aggressive (hitting, kicking, verbal put-downs)
   Self-care (dressing, washing)
   Routine-repetitive (eating, going to school)
   Constructive-productive (building, writing story, planning party)
   Physically exertive (sports, lifting heavy objects)
   Social-recreational (visiting someone, card games)
   Fantasy activity (doll play, cowboys and Indians)
   Directive (initiating, directing, demonstrating)
   Avoidance (stop trying, run away, shut eyes)
   Statement about self--positive, negative, neutral ("I have blue eyes." "I'm too stupid.")
   Problem-solving (producing idea, unusual combinations)
   Statements of information ("I know..."; non-evaluative observations about other people)
   Expression of emotion (crying, laughing)
   Conformity (express concern for rules, social norms, others' expectations, do as told)
   General verbal (trivial motor behavior such as dropping something, looking for something, listening)

4. Types of consequences:
   Positive consequences--
   From others--directed toward subject (praise, recognition, support, signs of affection)
   From self--self-praise, satisfaction
   From situation--reaching goal, unintended positive results
   Chance
   Author's statement, text
   Negative consequences--
   From others--directed toward subject (criticism, correction, rejection of ideas)
   From self
   From situation--inability to reach goal, unintended negative results
   Chance
   Author's statement, text
   Neutral consequences--not clearly positive or negative

*Saario, T.N., Jacklin, C.N. and Tittle, C.K. Sex Role Stereotyping in the Public Schools. Harvard Educational Review, 1973, Vol. 43, p.392-3.
©1973 by President and Fellows of Harvard College.

The "units" (items or reading passages) of analysis can vary and illustrations are included for analysis. It might be useful, particularly for the kindergarten and early reading tests, to analyze pictures on a separate tally form from the texts of items, since pictures are frequently used in the early tests.

A more extensive category system was used by Carolyn Jacklin in a study of early readers. Each character in each story was classified by age and sex and was coded on five additional categories: a. occurrence as a main character; b. occurrence in specific environment; c. occurrence as exhibiting specific behavior; d. occurrence as bearers of specific consequences; and e. occurrence as recipients of specific behaviors and consequences. The categories used by Jacklin are listed in Table 3.

* * *

Jacklin found satisfactory reliability, that is, consistency in agreement among raters in classifying characters in stories according to these categories. This elaborate classification scheme permitted Jacklin to construct several highly informative tables to summarize the data. Tables 4 and 5 are two illustrations of the summary tables that are possible.

* * *

Another form of summarizing the data was used for the environment category (in Table 6) and the same presentation was also used for types of behaviors performed by children and for types of consequences for children and adults.

Table 4

Total Number of Characters in the Sampled Stories Displayed by Person-Type*

(Saario, Jacklin, and Tittle, 93)

	Child	Adult	TOTAL
Female	241	124	365
Male	324	256	580
	565	480	945

Table 5

Number of Main Characters by Age and Sex*

(Saario, Jacklin, and Tittle, 93)

	Female	Male
Adults		
Number main characters	7	33
Total no. in stories	124	256
Chi square = 3.95; df = 1, p .05		
Children		
Number main characters	61	110
Total no. in stories	241	324
Chi square = 3.49; df = 1, p .05		

*Saario, T.N., Jacklin, C.N., and Tittle, C.K. Sex Role Stereotyping in the Public Schools. Harvard Educational Review, 1973, Vol. 43, p. 394 © 1973 by the President and Fellows of Harvard College.

Table 6

Types of Environments in Which Children (C) and Adults (A) of Each Sex (M/F), are Shown Given in Frequencies and in Percentages of Total Environments Shown by Each Age and Sex[+].

	Frequencies and Percentages			
Environment	CF n=241	CM n=324	AF n=124	AM n=256
Home	97 34.2%	111 29.0%	83*** 54.6%	59 23.6%
Outdoors	157 55.2%	234 61.1%	47*** 30.9%	144 57.6%
Business	15 5.3%	16 4.2%	8*** 5.3%	40 16.0%
School	15 5.3%	22 5.7%	14*** 9.2%	7 2.8%
Totals: Frequency	284	383	152	250
Percentage	100%	100%	100%	100%

* = p<.05
** = p<.01
*** = p<.001

[+]Saario, T.N., Jacklin, C.N., and Tittle, C.K. Sex Role Stereotyping in the Public Schools. Harvard Educational Review, 1973, Volume 43, page 396 © 1973 by the President and Fellows of Harvard College.

The use of a category system as extensive as the one developed by Jacklin has several advantages. The system can be used with curriculum materials as well as with tests and is detailed enough to permit discussion between teachers and students about sex bias and instructional materials, if the school desires.

The category systems used to analyze the content of educational tests have been less well defined but are also useful. The Donlon et al. study (22) classified the various roles attributed to females and males in the test content and also the relative status of male and female roles. Words that showed vocations, avocations, or special functions of people (for example, doctor, mother) were coded as roles. Roles were not inferred from the descriptions of individual behavior. For example, the role of "house husband" was not inferred from the sentence, "he cleaned the house and fixed dinner." The identification of particular roles as female, male, or neutral was decided by the percentages of females and males found actively engaged in that role as documented by the Occupational Characteristics, 1970 Census of Population. When the Census showed 80 percent or more of the individuals engaged in an occupation were one sex, the occupation was defined as a sex-typed role. Other occupations were classified as neutral roles. Historical consideration of roles was handled by general knowledge. If an item involved a role that was generally known as restricted to one sex (e.g., knights, the congressmen of 1800) it was coded as a sex-typed role. The assessment of relative status required both females and males to be present in an item. (The description of this

coding system is available in Donlon et al. 22.)

The coding system used for the study of the Metropolitan Achievement Test (MAT) content included an analysis of female and male illustrations as well. The analysis of illustrations categorized children and adults separately, according to three categories: male taller than female; female taller than male; female same as male. The classification of illustrations was in response to a criticism that boys are often shown as taller than girls, when in fact children of the same age are very similar in height regardless of sex. Also, the illustrations were examined for the portrayal of boys and girls and women and men in mixed-sex activities. Group activities was another category, i.e., single sex versus mixed sex groups, that may be useful for content analysis.

Four categories used in the "gender-balance" analysis of MAT written materials are presented in Table 7.

pronouns that are sex-linked have been used as another way of defining sex bias in educational achievement tests. Two types of systems have been used here and either seems appropriate for use by schools. In one type of analysis only the male and female nouns and pronouns are counted and categorized. In another approach to this count of language usage, the nouns and referent pronouns are counted for males and females and, in the case where sex cannot be assigned, a neutral category is tallied. A tally sheet would then have four categories: male nouns and pronouns; female nouns and pronouns; neutral nouns and pronouns; and neutral noun and pronoun referents. Donlon et al.(22) listed four identification procedures for tallying:

(1) The noun is inherently sex-linked, e.g., mother, father, sister, brother.

(2) The noun is found to have a sex-specific definition in the dictionary, e.g., ballerina: 1. a principal female dancer in a ballet company, 2. any female ballet dancer.

TABLE 7

Categories for Analysis of Gender-Balance (Jensen & Beck, 1978)

1. OCCUPATIONS
   Female Traditional: Nurse, Teacher, Librarian, Secretary, etc.
   Male Traditional: Laborer, Professional, Principal, Boss, etc.
   Female Non-traditional: Professional, Laborer, Boss, Principal, etc.
   Male Non-traditional: Nurse, Teacher, Secretary, etc.
2. ACTIVITIES
   Female Traditional: School, Playing with Dolls, Onlookers, Domestic Chores
   Male Traditional: School Sports, Games, other physical activities, adventurer, etc.
   Female Non-traditional: Sports, games, physical activity
   Male Non-traditional: Domestic Chores, Child rearing, etc.
3. ROLES
   Active: Main character, problem solver, giving help/gift
   Passive: Secondary character, needing help, recipient of help/gift
4. EMOTIONS
   Female Traditional: Fear, nurturance/tenderness, dependency, etc.
   Male Traditional: Aggression, courage, emotional strength, "strong silent type"
   Female, Male Non-Traditional: Cross-sex stereotypes

The other type of analysis is that of the frequency of usage of male and female nouns and pronouns in the test content. As noted earlier, the counts and ratios of frequencies of nouns and

(3) The noun is a definite female or male name, e.g., Bill, Mary.

(4) The noun has a female or male pronoun that refers to it, e.g., Pat went to her class.

In order to determine the relative balance of male to female references, the number of actors or nouns plus other words such as pronouns that refer to them in the item are counted. Repetitions are also counted.

Table 8 shows the summary table for the analysis of eight educational achievement tests reported in 1974 (Tittle, McCarthy, and Steckler, 112). This table has an extra column in it, since this study distinguished between generic nouns and pronouns (included in the category called All) and the category called Regular, which excluded the generic nouns and their pronoun referents. Table 8 lists a series of achievement tests and summary ratios of male nouns and pronouns to female nouns and pronouns; for example, the ratio in the California Achievement Tests, Level 3, grades 4 to 6, was 4:1, that is, four times as many male nouns and pronouns were used as female nouns and pronouns. The analysis by a school of an individual test would show the ratios for each individual subtest in the battery as well as the total set of items in the test.

This policy section has presented illustrations of the types of content and language usage analysis that teachers, counselors, and administrators may use to examine educational achievement tests for overt sex bias. There is another aspect to the analysis of content that should be mentioned. Earlier, a definition of content validity was given in terms of the match of the test to the curriculum. Content validity is a particular concern if the school is using new curriculum materials related to women's studies in history, social studies, and literature, for example. If a high school has included a section on the history of women in social studies or history courses then this should become a separate category for analysis in matching the test and local curriculum. Tests may not accurately reflect the changing content of courses in schools and this means paying special attention in reviewing the content of tests for their appropriateness in assessing local curricula.

If test content does not accurately reflect the local curriculum then test scores of students will reflect the mismatch. An example of the results of this type of effect was demonstrated in a study by Medley and Quirk (72) for the National Teacher Examination. In this study the effect of changing content specifications to include minority group history and cultural achievements was reflected in the relative performance of blacks and whites on the National Teacher Examination. Thus, the original definition of content validity is important as local curricula begin to reflect changing perspectives on women and their cultural and historical achievements.

The next issue is sex differences in performance on items. The section is particularly concerned with ways in which statistical evidence of item bias is developed.

ISSUE: ITEM BIAS AND CONSTRUCT VALIDITY

The issue with which statistical evidence of item bias is concerned is whether or not there are sex differences in performance on test items that do not appear to be related to the content and construct validity of the item. The main approach to providing statistical evidence is based on the percent passing the item for different groups, to determine whether there are some items on which males do better than females and, conversely, females do better than males. Plake, Hoover, and Loyd (84) provided an example of a test item in mathematics where females and males performed differently.

On the outside of the garage Mr. Nelson put a basketball goal 10 feet above the driveway. The goal was 2/3 the height of the garage. How many feet high was the garage?	
1. 6 2/3	3. 15
2. 13 1/3	4. (Not given)

TABLE 8

Ratios of Male Noun and Pronoun Referents to Female Noun and Pronoun Referents--Educational Achievement Test Batteries (Tittle, et al., 1974)

	TOTAL NO. OF TEST ITEMS	NOUNS AND PRONOUNS			
Test	*Items*	*All*		*Regular*	
		nM/nF=Ratio		*nM/nF=Ratio*	
California Achievement Tests					
Level 3 Form A (Gr. 4–6)	343	190/47	4.04	190/47	4.04
Level 4 Form A (Gr. 6–9)	337	84/46	1.83	84/46	1.83
Level 5 Form A (Gr. 9–12)	349	93/36	2.58	93/36	2.58
Comparative Guidance and Placement Program					
Form TPG (Gr. 13–14)	391	127/9	14.11	106/9	11.77
Form UPGX3 & UPGX4 (Gr. 13–14)	275	121/34	3.56	111/33	3.36
Iowa Tests of Basic Skills					
Form 6 (Gr. 3–8)	1232	1121/368	3.31	1211/368	3.29
The Iowa Tests of Educational Development					
Form Y5 (Gr. 9–12)	330	262/195	1.34	219/195	1.12
Metropolitan Achievement Tests					
Primary I Form F (Gr. 1.5–2.4)	174	51/59	.86	48/54	.89
Primary II Form F (Gr. 2.5–3.4)	257	137/86	1.59	137/86	1.59
Elementary Form F (Gr. 3.5–4.9)	300	124/42	2.95	121/42	2.88
Intermediate Form F (Gr. 5.0–6.9)	534	181/44	4.11	178/44	4.05
Advanced Form F (Gr. 7.0–9.5)	524	198/51	3.88	195/51	3.82
Sequential Tests of Educational Progress					
Series II Form 4A (Gr. 3–5)	420	366/103	3.55	322/98	3.29
Series II Form 3A (Gr. 6–9)	420	443/150	2.95	408/149	2.74
Series II Form 2A (Gr. 9–12)	470	468/134	3.49	360/120	3.00
Series II Form 1A (Gr. 13–14)	320	448/32	14.00	390/32	12.19
SRA Achievement Series					
Level 1–2 Form D (Gr. 1–2)	320	179/88	2.03	179/88	2.03
Level 2–4 Form D (Gr. 2–4)	276	333/241	1.38	330/234	1.41
Multilevel Form D (Gr. 4–9)	1070	1513/231	6.55	1462/229	6.38
Stanford Early School Achievement Test					
Level I (Gr. K–1)	126	217/93	2.33	217/93	2.33
Level II (Gr. 1)	259	192/168	1.14	190/168	1.13
Stanford Achievement Test					
Primary I Form W (Gr. 1.5–2)	251	134/53	2.52	123/51	2.41
Primary I Form X (Gr. 1.5–2)	251	119/78	1.53	115/78	1.47
Primary II Form W (Gr. 2–3)	409	209/89	2.34	192/87	2.20
Primary II Form X (Gr. 2–3)	409	143/87	1.64	143/87	1.64
Intermediate I Form W (Gr. 4–5)	540	221/83	2.66	198/71	2.78
Intermediate II Form W (Gr. 5–6)	544	171/58	2.95	166/58	2.96
Advanced Form W (Gr. 7–9)	532	181/46	3.93	157/46	3.41
High School Basic Battery Form X (Gr. 9–12)	478	245/50	6.13	242/39	6.21

The percent of males passing the item successfully was 55.4; the percent of females successfully passing the item was 37. One can speculate as to why an item such as this may show a difference in performance for girls and boys, but often there is no apparent reason for observed differences. There have been few studies that asked students to talk out loud during problem solving to identify what might be different approaches to a problem such as the one above. The item shown next was one in which the sex differences in performance were reversed, with more females answering correctly than males.

> In the 1968 Olympic games, the winner of the men's 200 meter dash was timed in 19.8 seconds. The winner of the women's 200 meter dash was timed in 22.5 seconds. The men's champion ran the 200 meter dash how many seconds faster?
>
> 1. 2.7 3. 3.7
>
> 2. 3.8 4. (Not given)

The percent of females answering the question correctly was 67.5 and the comparable percent for men was 55.8. From these examples it is clear that selection of items with different percents passing can affect the average scores of boys and girls on the total test score, since the total test score reflects the sum of the percent passing the individual items.

The percents passing the items, as illustrated above, are the data on which a number of different statistical approaches are based. The different approaches or methods that have been suggested for examining items by group (male-female) differences try to identify items on which there are statistically significant differences. The variety of techniques used all provide evidence on the construct validity of the items and test in a particular sense: reducing the relationship between the status variable of sex and item or test performance. The procedures all involve computing item difficulty data separately for males and females, and then using transformations, analysis of variance, chi square, or likelihood estimators to identify the "outliers" or items for which the two sexes are performing differently. These procedures are not detailed here but the interested reader will find various procedures described in Coffman, 11; Cardall and Coffman, 9; Angoff, 2; Angoff and Ford, 3; Echternach, 26; Schueneman, 99, 97; Fishbein, 31; Green, 38; Green and Draper, 39; Merz, 73; Veale and Forman, 114; and Merz and Rudner, 74.

Another series of studies have used analyses that control for ability level of groups. In general, the procedures cited above (with the exception of Schueneman, 99) do not attempt to provide a control for the initial ability of groups. Fishbein (32) also tried to take into account the ability level of groups in looking at whether or not they performed differently on an item in a particular test.

If groups of differing ability are used in studies, item difficulty indices change. This problem has led a number of researchers to suggest methods based on latent trait models, models in which the measure of an individual's ability is assumed to be independent of the distribution of abilities of examinees. Briefly, "A latent trait model specifies a relationship between observable examinee test performance and the unobservable traits or abilities assumed to underly performance on the test" (Hambleton and Cook, 41). Major concepts, assumptions, limitations, and examples of applications of latent trait models, including test bias, are given in a special issue of the Journal of Educational Measurement (57).

Latent trait models assume that only a single ability (or latent trait) is measured, and that the item responses of a given examinee are statistically independent. The item characteristic curve (icc) is a mathematical function that relates the probability of success on an item to the ability measured by the item set. The number of parameters required to descibe the icc depends on the particular latent trait model--the number of parameters is typically one, two, or three (Hambleton and Cook, 41).

The one-parameter model (the Rasch model) has been proposed and used to assess item bias by Durovic (24, 25) and Wright, Mead, and Draba (118), and Draba (23). The Rasch model assumes that all items in a set have equal discriminating power, and thus the items vary only in terms of

difficulty. As Hambleton and Cook note (41, p. 83), this is a very restrictive assumption and likely to be violated with most tests. On the other hand, the three-parameter model requires large numbers of examinees to estimate parameters and extensive computer time (Lord, 65).

Ironson (52) has compared four of the procedures proposed: the transformed item difficulty (Angoff), the item discrimination (point biserial), the chi square (Schueneman), and the item characteristic curve (three-parameter model) method. The methods identify different items as biased, in her study, and show low to moderate correlations between methods. The item discrimination method was least related to the other three models.

While there are no studies on which to base consensus or expert agreement of the extent to which the use of one method or another is preferable, it appears that the use of at least one of these procedures, with a clear specification of the decision-making rules for including items in the test, would enable the test user to make a judgment as to whether or not the test is fair for a particular group of girls or women. The absence of such a procedure appears to be more important than the use of a particular procedure, until optimal procedures are identified.

In terms of assembling the total test, for norm-referenced achievement tests particularly, only Diamond (20) has suggested a decision rule for selecting items. Diamond would reduce mean differences between subgroups by examining subgroup differences when selecting items on the basis of pretest data. Differences between the percentage of upper and lower groups selecting the correct answer can be calculated separately for each subgroup. The difference between subgroups should not average 5 percent in either direction for any given subtest, according to Diamond. Other recommendations for dealing with this part of the test development process have not been located. The decision rules used at this particular stage of test construction are another procedure that test users should look for in test manuals, and the use of these rules will help to define whether a test is fair to females and to males.

Another type of item bias may be more generalized over a set of items. In some instances it may be possible to identify a particular item response format or set of directions that have a differential effect for females and males. A study of one particular item alternative, "I don't know" as a response, was carried out by Sherman (100). She examined the use of the "I don't know" alternative in the National Assessment of Educational Progress. In the beginning of each NAEP administration, respondents were instructed how to answer the excercises and were shown a sample multiple-choice excercise. A tape recording was played during each administration. The following instructions concerning the uncertaintly or "I don't know" alternative were read: "If you don't know the answer to an excercise, just fill in the oval next to I don't know." (Sherman, 100, p.2). After each multiple-choice exercise was read to the respondents, the announcer added, "If you do not know the answer, please mark the 'I don't know'response." (p.2) Sherman's idea was that respondents choosing the "I don't know" response could be responding on the basis of uncertainty but also on the basis of personality variables, that some individuals might be less self confident, rather than less knowledgeable. She hypothesized that this relationship might vary according to group membership. Testing her hypothesis, Sherman found that an adjustment for the use of the "I don't know" response by groups had a large impact on the sex differences in science performance. Sex differences at the three younger ages tested by the NAEP were reduced by a regression analysis modification. The sex differences were virtually eliminated at the adult level over the 66 excercises analyzed. As Sherman noted, "Sex differences in correct response percentages for many of the excercises at the adult level can be explained almost completely by differences in usage of the "I don't know" alternative. Some exercises continued to show a clear advantage for one sex or another after the data modification but there are fewer showing the overwhelming male advantage as depicted in NAEP data." (100, p.14).

The particular form of sex bias detected by Sherman would probably not be identified in the analysis of a set of items by the types of procedures described above, where sex differences in performance for

individual items are examined. This is a particular response format that would affect all the items in a test, providing a consistent effect across all items. The item bias procedures developed so far cannot identify such a difference. Teachers and administrators should be alert to considering a variety of hypotheses for why one group may score differently than another where there is no likely educational reason for group differences to appear. The construct validity of the achievement test is poor if response format or test directions affect groups differently. Another situation, not necessarily applicable to male or female performance, but perhaps to groups who more or less frequently take tests, would be frequent changes in the manner in which test items are presented. Students who are more used to taking tests might tend to score higher on such a test than students who are not used to taking tests and responding to a variety of different directions. (See Schueneman, 98, for further examples of the effect of item ordinal position and item format on subgroup differences in performance.)

Policy

The research concerned with the issue of statistical evidence for item bias suggests that teachers and school administrators need to formulate policy in relation to the review and selection of achievement tests for use in schools.

There are several types of data that reviewers of educational achievement tests should find in test manuals. The most critical data will be evidence that item analysis procedures have been carried out to examine whether females and males performed differently on indificual items in the test. The second type of evidence is a description of the decisions used to include or exclude items where females and males perform differently in try-out or pilot testing of items. There should also be a list of the items and the percents of females and males passing each item for all items included in the final version of the test. This listing of the percent passing the items should also show the frequency distribution of item percent passing separately for females and males.

Within content specification limitations, adjustments in the items selected should reflect roughtly comparable item difficulty or percent passings for females and males to minimize differences between the two groups in the final test distribution for comparable sub-groups. The procedures and the resulting data should be reported in a test manual to provide the test user with the evidence to make the decision that the test is a fair test for females and males. However, despite the use of these procedures in item analysis and item selection, sex differences in average performance may remain.

ISSUE: TEST BIAS AND CONSTRUCT VALIDITY

Where sex differences in average performance on educational achievement tests remain, the test publisher and the test user need to seek evidence of construct validity. That is, does the test measure the same construct for each group or is the test "biased" in the sense of measuring different constructs for each group?

As mentioned earlier, construct validity is needed since test users often make inferences beyond the actual content of the test items. Cole (12) suggests that a test interpreter rarely, if ever, remembers the particular types of items on a test when interpreting children's scores. On the score report, there will only be the label "math concepts" or "science content" or "science process." She reports that on the basis on content categorizations of items on the science achievement test of the International Association for the Evaluation of Educational Achievement (IEA) sex differences in science achievement have been reported: girls perform more poorly than boys on physics items and on "understanding" items. The construct interpretation that users tend to make is that boys achieved higher levels of understanding of science than girls, and sometimes the inference extends to the additional judgment that girls do not have the capability for high levels of science. Cole reports a study by Carlson demonstrating that at least a few of the "understanding" items involved sex-differentiated practical experience. Boys did better on an "understanding" item involving how to put batteries in a flashlight. Girls did better on an "understanding" item about how to place a jar under hot water to get the lid off. Wherever sex differences remain in performance on educational

achievement tests, publishers need to demonstrate, and test users in schools should require, evidence of construct validity. One type of evidence of construct validity is to experimentally manipulate item content.

A study by McCarthy (71) experimentally manipulated the context of mathematics items and showed that girls performed better on items that had a context of "typical" activities of women in our culture. Her study suggests that it is possible to construct item pools and develop a test on which we expect similar average performance of females and males in mathematics or science achievement. Where performance differences remain for groups of females and males, the test publisher needs to examine other plausible rival hypotheses that may account for differences if the test is claimed to be fair to both sexes.

A second type of study is a demonstration within subgroups of females and males that similar patterns of correlations exist for the achievement test with a specified set of other measures or criteria (such as aptitude measures). Ironson's study (52) suggests that this approach of correlating achievement and aptitude measures to examine the patterns of intercorrelations separately by sex may be feasible, since she found fewer biased items in more abstract, less achievement-oriented measures (e.g., picture-number and letter groups). As noted above, the few studies of test construct validity and sex bias have examined tests of mathematics and have varied the familiarity of the context in which the mathematical process was embedded. There is a particularly critical need to examine the construct validity of tests at the high school level in the areas of science and mathematics, areas that women are not traditionally encouraged to view as "feminine occupations" or fields of study.

Another approach can be tried, similar to the System of Multicultural Pluralistic Assessment (SOMPA) Mercer has developed for the WISC. Publishers can develop measures descriptive of each individual's relevant experience in mathematics, as well as attitudes toward mathematics, as another approach to determining construct validity for both females and males. Hypotheses can be stated for the predicted relationships among correlations with achievement test scores, past "experience" in mathematics, and attitudes toward mathematics. Specification and provision of evidence on these variables might give the user a "frame of reference" of other variables to determine that a test is sex-fair to individuals even though average differences in group performance may remain.

Policy

If the test user is given evidence to determine that sex differences in test performance still remain after item bias studies, careful item selection procedures, and studies of construct validity, what policies can be recommended for test interpretation and counseling?

Two issues can be briefly noted here and placed in the context of policy decisions. The first is the issue of whether or not there should be separate sex norms, combined sex norms, or, perhaps, experienced-reference norms if mean sex differences in performance remain on educational achievement tests. The second issue is the relationship of test performance and interpretation to course selection.

Achievement tests do not now provide separate norms since there are few areas of achievement showing sex differences. In science and mathematics, key areas of concern for occupational desegregation, there may be sex differences in performance at the high school or college level. In these instances, the best policy may be to continue with combined sex norms and to provide inservice training to teachers and counselors, and additional interpretative materials to students and parents, information that may help to place test scores in context. We can reasonably assume that we do not know whether any sex differences remaining are due to social or innate factors, and that in any event for educational purposes it is more logical to assume that they are the products of cultural experiences, including sex role socialization practices.

Workshops for teachers and counselors can focus on test score interpretation for individual students, pointing out the stereotypes sometimes held for female and male performance. Awareness of the importance of basic areas for later career choice, especially in mathematics and science, can be stressed. School adminis-

trators can examine sex differences in achievement test scores and course enrollments within their own schools. At the high school level, are there sex differences in math and science achievement for boys and girls? If there are differences in performance, are there indicators that attitudes of the staff differ for boys and girls? Are different courses counseled for boys and girls? Are staff classroom behaviors equally responsive to the performance of boys and girls, or with girls and mathematics particularly, are there specific steps taken in instructional design to ensure that girls succeed and develop a positive attitude toward mathematics to the same extent the boys do?

Assertive, affirmative, action is needed for many girls who will not consider additional mathematics courses without suggestions and support from teachers and counselors. The development of special college courses to assist women to overcome fears of further education in mathematics is ample evidence of the need to work more affirmatively in the junior and senior high schools.

Teachers and counselors must show the same assertive action in assisting girls to continue work in the sciences. Courses in physics may be open to girls, but more is required: girls must be actively recruited to pursue course work in the sciences and technical fields. Teachers and guidance staff can relate achievement test areas to occupational clusters and emphasize the basic importance of mathematics and science in leaving one's "options" open. Stereotypes and prejudices toward women in math, science, and the trade/technical areas can be explicitly addressed. Confronting the folkways and prejudices can assist boys and girls to examine their own attitudes toward the achievement of each sex in these areas.

Teachers and counselors can also work with students and parents in reviewing test manuals and achievement tests for evidence of sex bias, to examine interpretive materials for overt sexism, and to review their own interpretive actions with respect to individuals for sexism. Parents can be assisted to examine their own values in relationship to student achievement scores. Again, the key areas for exploration here are mathematics, science, and nontraditional occupations for both sexes.

POLICY SUMMARY: EDUCATIONAL ACHIEVEMENT TESTS

Sex bias in educational achievement testing can be examined at several levels. First, there is the review for overt sexism. Second, item analysis data and decision rules for selecting items can be examined in manuals provided by test publishers. Where mean sex differences in performance remain in educational achievement tests, review panels selecting tests, whether at local schools or for statewide assessment, need to determine whether the publisher has provided evidence on the construct validity of the achievement tests. Materials that accompany the tests, test administration directions, all manuals, forms, and other interpretive materials accompanying the educational achievement test also need to be examined for overt sex bias and for their assistance to students, parents, and teachers in understanding the possible effects of sex role socialization on student performance. In-service courses for teachers, counselors, and administrators will help to insure sex-fairness in presentation and interpretation of achievement test scores. All of these actions are directed toward assisting girls to experience success and eliminate sex stereotypes of achievement in any area, but particularly in mathematics and science, since these are critical areas influencing later career choices and options. Career interest measures are also influences on occupational choices.

SEX BIAS AND CAREER-INTEREST MEASUREMENT

OVERVIEW: OBJECTIVITY AND SUBJECTIVITY IN MEASUREMENT

Tests that assess an individual's interest in different occupational areas or basic fields of interest comprise a small percentage of tests sold on an annual basis. However, interest assessments are included in several major college admission programs and it is estimated that 3.5 million interest tests are administered annually (111) including those given as part of college admissions testing.

The interest measurement area provides a study in the nature of scientific psychology and whether there is "objectivity" or "subjectivity" in its empiricism. Two of the major interest inventories, the Strong Vocational Interest Blanks for men and women and the Kuder Occupational Interest Survey DD (KOIS) were empirically developed. Both of these measures, until very recently, reported a separate set of occupational scales to females and males. The scales reported for women were based on a smaller number of occupations and emphasized traditional women's occupations.

The inventories were developed on an empirical basis, with little in the way of a theoretical formulation to guide them. Theoretical formulation in this case refers to a broader conception of the domain of occupations, career guidance, and the fundamental assumptions made about the occupational world and its relationship to males and females. If equal opportunity and rights are of basic concern, the stricly empirical basis of developing vocational interest inventories was not adequate to meet these concerns. There was not a consideration of alternative ways to assess interests and to ensure that women were encouraged to consider choosing occupations not typically thought to be open to them. Both the Kuder and the Strong made it clear to women that there were separate occupational scales for men and women. There were 77 occupational scales for men and a total of 57 for women (which included 20 on the men's scale). There were 17 scales that overlapped and appeared in both the female and male occupational norms. The KOIS also presented college major scales for females and males. These scales were particularly discriminatory, since college major scales for women did not include the following areas: law, business management, marketing, finance, government, pre-medical, dentistry, and pharmacy. These items are noted as background, to give the status of interest inventories up to the early 70s.

In the 1970s pressures from the professional guidance organizations and the National Institute of Education project which formulated Guidelines on Sex Bias and Sex Fairness in Career Interest Measurement were instrumental in encouraging some change. The major changes are noted below, but, as described in the policy sections, the changes are incomplete thus far. And the changes in tests are only part of the changes required in the counseling system within which they are used.

Types of Interest Inventories

There are two types of scale construction traditionally used for interest inventories. Occupational scales are based on the empirically determined relationship between the interest expressed by the taker of the interest inventory and those of individuals already employed in occupations. Two examples of this type of scale are the occupational scales of the Strong-Campbell Interest Inventory (SCII) and those of the Kuder Occupational Interest Survey DD (KOIS). Homogeneous scales, scales based on internal criteria, are developed through some form of clustering items -- similar types of activities for a job or interest areas, the theory of the test constructor, perhaps sorting by judges, or intercorrelation of items and

scores. The responses of the test taker are reported for scales internal to the instrument. Example of instruments with homogeneous scales are Holland's Self-Directed Search (SDS), the American College of Testing Program Interest Inventory (ACT), and the Ohio Vocational Interest Survey (OVIS).

The two types of scale construction have led to different problems in examining sex bias. These problems are reflected in the NIE Guidelines for Assessment of Sex Bias and Sex-Fairness in Career-Interest Inventories (Diamond, 19). These guidelines are summarized here and the major issues are discussed in more detail for particular guidelines.

The NIE Guidelines

As Diamond (19) described the process, the working definition of sex bias used in the development of the NIE Guidelines was:

> Within the context of career guidance, sex bias is defined as any factor that might influence a person to limit--or might cause others to limit--his or her consideration of a career solely on the basis of gender.

The NIE Guidelines have three major areas: the inventory itself, technical information, and the interpretive information that accompanies the inventory. The first area, the inventory itself, is concerned with the "face validity" of a sex-fair instrument--overt sex bias and sex-role stereotyping. The technical information section includes issues specific to the homogeneous scales and the occupational scales. The interpretive information section is concerned with materials for students and counselors, to reduce sex-role stereotyping of occupations and to encourage career exploration activities. The NIE Guidelines are briefly summarized as follows.[3]

(1) The inventory itself.

a. The same form should be used for women and men unless it is empirically shown that separate forms minimize bias.

b. Scores should be given on all occupations and interest areas for both women and men.

c. Item pools at the inventory and scale levels should reflect experiences and activities equally familiar to each sex.

d. Occupational titles should be present.

e. Use of the generic "he" should be eliminated.

(2) Technical information.

a. Technical manuals should describe how the inventory meets these guidelines.

b. The rationale for separate scales should be given.

c. The same vocational areas should be indicated for each sex even if it is empirically demonstrated that separate inventory forms are more effective in minimizing sex bias.

d. Sex composition of the criterion and norm groups should be described.

e. Criterion and norm data should be updated every five years.

f. The information on career options distributions suggested for each sex should be provided.

g. The validity of interest inventories for minority groups should be investigated.

---

[3] Stebbins, Ames, and Rhodes (102) provide extended comments on the rationale for each guideline and give examples of their purposes. Diamond (19) is a useful reference for the technical and research issues, as well as the research studies in Tittle and Zytowski (111).

(3) Interpretive information.

a. Interpretive materials should point out that vocational interests and choices of women and men are influenced by many environmental and cultural factors, including early socialization, sex-role expectations, and home-versus-career conflict.

b. Orientation to the inventory should encourage respondents to examine stereotypic sets toward activities and occupations.

c. The users' manual should state that all jobs are appropriate for qualified persons of either sex and should attempt to dispel myths about women and men based on sex-role stereotypes.

d. Interpretive materials should encourage exploratory experiences in areas where interests have not had a chance to develop.

e. Case studies and examples should represent men and women equally and include examples of each in nonstereotypic roles.

A majority of guidelines apply to interest inventories and accompanying manuals regardless of whether the inventories have occupational scales or homogeneous scales. Individual issues and policy recommendations follow.

ISSUE: FACE VALIDITY

"Face" validity is concerned with overt bias in occupational titles and sex-role stereotyping in interest inventories. Guidelines 1a, d, and e and guidelines 3a, b, c, and e are concerned with this issue. With most interest inventories, it is now possible to use the same form of the interest inventory for women and men. One exception is the Minnesota Vocational Interest Inventory (MVII). The MVII has been developed for occupations in the non-professional fields and is based entirely entirely on the interests of men in these occupations. Under Title IX, this inventory could not be used unless a similar inventory were available for women. Its use in assisting only male students for vocational education program placements would be discriminatory against women. The publisher to date has not provided a comparable form for women, nor has there been an attempt to revise or develop scales for the MVII. The administration of the MVII would also be prohibited under the definition of sex discrimination used in the Vocational Education Act of 1976.

Policy

The most important policy for school administrators and counselors is the decision to develop a review procedure and a checklist form for career-interest measurement and accompanying guidance materials, derived from the guidelines concerned with the "face validity" of instruments and materials. The checklists should include the following items: Is the same test form (set of items) used for both girls and boys? Earlier forms of tests that had separate forms for men and women should be discarded. Are all occupational titles in gender-neutral terms? Terms that restricted occupations to males or females should be restated in such form as, for example, firefighter, letter carrier, and flight attendant. Is the generic "he" eliminated from all test and interpretive materials?

For interpretive information, the guidelines in 3a, b, c, and e should be in the checklist for review: Are interpretive materials provided that describe the possible effects of early socialization, sex-role expectations, and home-versus-career conflict? How are the issues of cultural factors that influence the vocational interests and choices in women and men included? Are students and counselors encouraged to examine sex-role stereotyped ideas about the appropriate activities and occupations for females and males? Does the manual and the student interpretive material state that all jobs are appropriate for qualified persons of either sex? Are there facts presented about women in the world of work that dispel myths about women? In sum, the instrument and its interpretive materials must take a positive, strong affirmative action position.

ISSUE: SAME SCALES FOR FEMALES AND MALES

Guideline 1b, scores should be given on all occupations for both women and men, applies particularly to the interest inventories with occupational scales. Both the SCII and the KOIS, for example, use the same set of items for the basic inventory, yet the criterion or occupational groups are constructed separately by sex. The issue is that women still know that the occupational world is differentiated by sex, since scales are normed or developed separately for females and males. On the SCII there is an occupational scale for physician (m) and physician (f) on the profile form for test results. These scales are constructed on the basis of responses of a group of male and female physicians, respectively. Response differences between the occupational group and the men- or women-in-general groups are used to derive the scale scores. The profile reports both scores, asterisking the same-sex score (i.e., an asterisk appears by the physician (f) for a girl's profile). This practice reinforces the stereotype of sex differences in occupations and is presently justified only by item data that show sex differences in response rates. For example, items such as the following show differences (Stebbins, Ames, and Rhodes, 102): "Would you like to race automobiles?"

criterion group from a sample of men-in-general or women-in-general. In the example in Table 9 the responses of male race drivers are not very different from those of a representative sample of men (men-in-general). When a male responds "like" to the item, there is little here to distinguish him from male "race drivers." However, when a woman responds positively, she has given an unusual response, one that differentiates her from women-in-general and that is similar to female "race drivers." This type of item response means that the item would appear on the female-normed scale for race driver but the item would not be used for the male-normed scale for race driver.[4]

The KOIS does not use men-in-general and women-in-general groups in constructing occupational scales. Differences in occupational scales for men and women are based on the response differences between men and women in the same occupation. Scores on cross-sex scales for the KOIS (same-named occupational scales for females and males) therefore reflect directly any sex differences in response. On the SCII however, with the use of the in-general groups for scale construction, the more "traditional" a woman's interests are, the less likely she is to

TABLE 9

Response Rates for an SCII Item

Response	*Female race drivers*	*Male race drivers*
Like	95%	95%
Indifferent	5%	5%
Dislike	0%	0%
	Women-in-general	*Men-in-general*
Like	10%	75%
Indifferent	10%	20%
Dislike	80%	5%

Occupational scales on the SCII are constructed by examining the differences in responses for members of an occupational

respond the way the men-in-general group does, and therefore the higher her score on male-normed scales is likely to be

[4] Johansson and Harmon (55), Campbell (8), Hansen (42), Webber and Harmon (116) provide further example of analyzing the SCII for sex differences and the implications for occupational scale construction.

(for traditionally female occupations). Conversely, on the KOIS, women will often get lower (Lambda) scores on cross-sex norms than on their own (same sex) norms.[5]

The items currently used in instruments with occupational scales were not examined for sex bias when originally constructed. That is, the items were not selected to be "equally desirable" by a general sample of both females and males. In the absence of this criterion for item selection, the development of new occupational scales is limited to items that are already in the item pool and that vary on this characteristic. As one example, Johansson and Harmon (55) examined items common to the earlier Strong Vocational Interest Blank (SVIB) for females and males--two separate forms--for 14 occupational criterion groups. They found between 15 percent to 21 percent of these items represented sex-stereotypic responses only and hence were not valid in differentiating occupationally related differences.

A study by Harmon and Conroe (46) examined the sex stereotyping of occupational titles versus occupational activities. Interest inventories differ in their use of the titles or activities of occupations in items. The Harmon and Conroe study provides some evidence that occupational titles are perceived more stereotypically than activities. That is, "doing research work" may not be perceived as stereotypically as "scientific research worker."

As these studies suggest, the ultimate goal for the interest inventories is a set of scales that do not incorporate sex differences. This goal is stated in guideline 1c, item pools should reflect experiences and activities equally familiar to each sex, and is not presently met for the SCII and other measures using occupational scales. (The guideline has been met with some homogeneous scales described below). As a result, the test user must carefully examine the interpretive materials and the guidance setting for steps taken to counter the sex stereotyping of occupations that remains in the instruments. The different norms (male and female) can be used positively to initiate discussion on sex-stereotyping of occupations and activities if counselors so desire.

Guideline 1c, however, has been largely met with the item pool for one of the interest inventories with homogeneous scales--an experimental form of the ACT, the Unisex Interest Inventory (Uni II). Rayman (90) examine the ACT IV items for sex differences, had sex-balanced items written (i.e., items likely to exhibit 10 percent or less difference in "like" responses between the sexes), and constructed the Uni II. This instrument was administered along with the ACT IV to 3,000 college-bound students. The results showed the average differences between females and males (percentage of "like" responses) were much smaller for the Uni II than for the ACT IV. There were no significant differences between raw score means for the sexes in the Realistic, Artistic, and Conventional Scales; statistical (but not practical) differences were found for the Investigative and Enterprising Scales. The sex balance was least well-achieved for the Social Scale. Hanson and Rayman (44) pursued this development and examined validity-related analyses. They conclued that the sex-balanced scales were generally valid, as measured against the ACT IV.

Policy

Since scores on all occupational scales are currently reported to all users, the policy question focuses on interpretation. How should counselors and students work with all the scores presented to students? Similarly for homogeneous scales, how should counselors and students view the same-sex and combined-sex (or opposite sex) norms reported. Special interpretive material is required, and in-service workshops are needed for teachers and counselors to explore the differences between the female and male scales and norms reported to individuals.

Administrators and counselors should carefully review the reports received by students for the inclusion of discussion of the two sets of norms, of the likely

[5] Johnson (56) and Lunneborg (66) provide further examples and discussion of the issue of cross-sex norm interpretation. Johnson has recommended the elimination of the men-in-general and women-in-general groups in the construction of SCII scales. Tittle and Denker (130) show this effect on the KOIS.

influence of sex-role socialization on interests, and their relationship for the particular boy or girl receiving the results. Both written materials and intereactions with counselors are the best policy for schools.

ISSUE: CAREER OPTIONS SUGGESTED TO FEMALES AND MALES

Guideline 2f is concerned with another issue of the homogeneous scales--the distribution of career options suggested for each sex. For example, are careers in the interest area of science suggested in the same proportion (overall frequency) to both boys and girls? Ideally, according to one definition of sex-fair testing[6], the proportion should be the same. Hanson and Rayman have demonstrated that the same raw score means can be achieved using sex-balanced items. However, with the usual unbalanced item pools, this equality does not occur and therefore raw scores are "normed" (adjusted to within-sex means and standard deviations, and placed on a scale with the same standard deviation for each interest area). This procedure highlights another controversy since unequal distributions of suggested options result for the sexes unless same-sex norms are used.

Hanson, Noeth, and Prediger (45) examined four ways of reporting scores: interest profiles based on (a) raw scores, (b) combine-sex norms, (c) same-sex norms, or (d) opposite-sex norms. Their samples were tested with either the VIP or ACT (1970 or 1972) and followed up in college in 1975. They concluded that same-sex norms provided results showing criterion-related validity as high as or higher than the other procedures, and same-sex norms offered the additional advantage of suggesting similar vocational options to females and males (guideline 2 f). (See also Prediger and Hanson, 86, for similar conclusions.) Where raw score distributions for interest areas are not the same for females and males, same-sex norms are required to achieve this goal.

Gottfredson and Holland (36) presented a study in which the use of the SDS raw scores appeared to be more efficient predictors of self-reported occupational choices (1- to 3-year follow-up) than sex-specific norms. Holland (49) has pointed out that interest inventories have two pruposes--exploration and prediction--and that some may function better than others for each purpose, as part of the justification for using raw scores with the SDS. Prediger and Cole (85) argued strongly for a different criterion and examined the relationship between sex-role socialization, employment data, and interest inventories.

Policy

Although not specifically taken up as an issue in the NIE Guidelines (except in guideline 2f), this issue may be resolved in policy and practice by requiring that both types of information are reported to the student in vocational guidance. That is, if both the same-sex and opposite-sex norms (or both raw scores and normed scores) on any homogeneous scales are reported, students can question and explore whether response differences are due to sex-role socialization and limited experience or to genuine differences in their interests as individuals. Recent recommendations of the Office of Civil Rights (OCR) and the Association for Measurement and Evaluation in Guidance (AMEG) Sex Bias Commission include providing scores on both sets of norms for everyone. This permits individuals to determine how they rank with others of their own sex exposed to similar socialization experiences and how they rank with individuals of the opposite sex.[7]

---

[6] Prediger and Hanson (in Diamond, 19) defined sex restrictiveness as evidence of sex bias: An interest inventory is sex-restrictive to the degree that the distribution of career options suggested to males and females as a result of the application of scoring or interpretation used or advocated by the publisher is not equivalent for the two sexes. Conversely, an interest inventory is not sex-restrictive if each career option covered by the inventory is suggested to similar proportions of males and females. A sex-restrictive inventory can be considered to be sex-biased unless the publisher demonstrates that sex-restrictiveness is a necessary concomitant of validity.

[7] E.E. Diamond, personal communication, February 1978.

For the SDS, which reports only raw scores for each scale, there is no easy path to a judgment of equal access to occupational programs. The distributuions of suggested areas for career exploration do differ by sex (reflecting the current occupational segregation patterns). The Holland scales can be interpreted as reflecting socialization practices, and women tend to score highest on the Social and Artistic scales and very low on the Realistic scale. The pattern of sex differences does not meet definitions of sex-fairness in counseling. A sex-fair policy would be to provide extensive supplementary interpretive materials for the student and career guidance staff; the SDS should not be used by itself in sex-fair career guidance and counseling.

ISSUE AND POLICY: MINORITY AND MATURE WOMEN AND INTEREST INVENTORIES

The last technical guideline is one for which limited information exists: guideline 2g recommends the investigation of the validity of interest inventories for minority groups. Gump and Rivers (40) reviewed the litereture for validity studies for black women and found little direct evidence of validity.[8] They concluded that the validity of current inventories, with occupational scales constructed on predominately white samples, was questionable. Lamb (62), however, studied the validity of the ACT Interest Inventory for classifying students into educational major groups, for both females and males of five ethnic groups. Lamb found the structure of interests was comparable across the white and minority samples, with the exception of Native American (Indian) males. The accuracy of classification was comparable for most minority groups. Lamb's sample consisted of college seniors, however, and it is not clear to what extent these findings would generalize to the secondary level. Administrators and counselors need to look to the interpretive materials and career guidance activities to make a determination of sex-fairness for minority women, since technical data to support sex-fairness for minority group women are not available. As appropriate, counselors need to consider whether culture-specific experiences will limit women's opportunities.

Similar cautions apply to the use of interest inventories with mature women. Although there is some evidence (Denker and Tittle, 17) that older women view interest inventory results as appropriate, there has not been systematic work looking at item responses for this group of women. Again, counselors will need to use interest inventories to explore options, not close them.

ISSUE: THE PURPOSE OF INTEREST INVENTORIES AND VALIDITY

An issue of concern is the type of validity emphasized for interest inventories. Holland (49), as mentioned earlier, pointed out that interest inventories may have two purposes--exploration and prediction. Prediction is based on a model of human behavior that says from one point of time we can predict a future event, such as occupational entry and satisfaction. Implicit in this model for prediction is the assumption that there is stability in behavior. For women, this is the wrong assumption at present. The current emphasis in intervention programs and counseling is to assist women to explore other than traditional views of occupations. This emphasis is likely to insure a more "unstable" or less predictable set of patterns of career choices for a period of time. Taking this point of view, it is more appropriate to emphasize what may be called the exploration validity of interest inventories, rather than predictive validity. The effect of interest inventories should be to provide satisfactory exploration validity. This term is not well defined yet (Super and Hall, 105) but can be tentatively defined as the extent to which the career interest inventory is useful in stimulating the student to explore--to seek information about occupations new to the student and to try new activities that may be related to career choice. The effects of career-interest inventories

---

[8] Picou and Campbell contains a series of papers on minority groups but again lacks data on interest inventories.

focused upon here, along with any interpretive materials or guidance activities provided with the inventory results, are stimuli to exploration. (Guideline 3d suggests that interpretive material should encourage exploratory experiences in areas where interests have not had a chance to develop.)

Recent research results suggest that counselors should focus on these effects, and that test publishers and career guidance staff should be designing interpretive material for interest inventory results that encourage more career exploration. Holland (48, 49) and Holland, Takai, Gottfredson, and Hanau (50) have summarized exploration-related validity studies (using experimental designs and the SDS) for the number and types of occupations considered by high school level clients that support the use of the SDS for exploration. The Holland et al. (50) study used high school girls only (N=252).

Cooper (14) compared other materials, using a control group design. She examined the effects of the SCII, a non-sexist Vocational Card Sort (VCS) (Dewey, 18) and auxiliary materials designed to make women respondents aware of myths and realities of women in the world of work. For her sample of college women a few differences were found. The VCS was more effective in broadening career options and increasing the frequency with which women students read occupational information.

Zytowski (119) assessed the effects of giving KOIS results to high school juniors and seniors (N=100). Both boys and girls were in the sample. There was an increase in accuracy of self-ranks on expected order of scoring on a set of selected occupations. There was no increase in self-reports of information-seeking behavior as a result of receiving KOIS profiles and interpretation, in contrast to studies of the SDS.

A study by Prediger and McClure (87) used an intervention with ninth and twelfth grade high school girls to try to encourage exploration of careers in science and technology. The interventions included, for ninth graders, a career interest inventory (ACT-VIP) and group discussions of career planning. The experimental group reported more career exploration activity (but not specific to science and technology). The twelfth grade college-bound girls were mailed a booklet on careers for women in science and technology designed to encourage women through role models and discussion of myths on women's careers. The intervention was not successful in terms of the outcome measures used.

Takai and Holland (106) have also examined the effects of the VCS, the SDS, and the Vocational Exploration and Insight Kit (VEIK), a combination of the VCS, the SDS, and a plan for exploration activities. In a sample of 241 high school girls there were no significant differences between the effects of the VCS and the SDS on the variety-of-occupations criterion, and the SDS was more effective than the combination treatment of VEIK. No significant differences were obtained on a satisfaction scale criterion. Other criteria examined but not used in the analysis were pre-post measures on the number of occupations considered, satisfaction with choice, and the variety of information-seeking activities (self-reported). A no-treatment group was not included in the design. Oliver (81) summarized research on modes of test interpretation and listed additional criteria such as accuracy of self knowledge, certainty of choice, and realism of choice. These criteria may be useful in addition to the exploration-related criteria (also reported in Oliver's study).

In these series of studies different criteria for exploration validity have been used. Counselors and administrators may find it helpful to determine which ones are most important as outcomes for students in their schools.

Policy

There are several policy implications for career guidance based on the studies of the effects of interest inventories in increasing career exploration. The first is that this changing emphasis in career guidance should be considered in order to determine its relevance in the local school. If career exploration is seen as the primary goal in using interest inventories and other career materials, administrators and counselors need to determine their own criteria for this goal. Counselors may want to establish procedures' checklists of the activities, in addition to interest inventories, that are available to students to promote career exploration. These checklists would assist students and counselors to know the resources available locally to find informa-

tion on careers, to talk to resource teachers or persons outside the school, and to have temporary, short-term work or site visits to provide better understanding of the nature of specific occupations. These same checklists can be used to provide criteria for determining the effectiveness of local efforts to increase career exploration for both girls and boys. The frequency of student use of various activities, when records are kept, will provide evaluation for the most effective parts of programs or will serve as a stimulus to try to increase student use of particular types of career exploration activities.

The studies thus far on the effects of interest inventories suggest that on their own, without supplementary material, they are not as effective in stimulating exploration as they might be. The research studies are also limited, with the exception of Zytowski's (119) study, in focusing more recently on samples of girls only, so that the comparative effectiveness of exploration-designed activities for females and males cannot be examined. And the effectvieness of interest inventories and activities in encouraging exploration of non-traditional occupations for both girls and boys has not been determined.

ISSUE: COUNSELOR USE OF INTEREST INVENTORIES

One of the issues in sex bias and interest inventories has been a concern with the counselor's use of interest inventories and possible counselor bias in counseling females and males. Most of the research on sex bias and interest inventories has examined the inventories themselves, their technical development, and interpretive materials. To date, there have not been studies that examined whether counselors interpreted interest inventory results differently according to a client's sex. Oliver (81) evaluated career counseling outcomes for three types of test interpretation. However, there was no analysis of possible differential treatment effects by sex of the client.

There are related studies of counselor attitudes, however. These studies have examined the question of whether knowledge of a person's sex will affect the educational and occupational expectations, evaluation, or treatments provided by a counselor. Harway and Astin (47) have reviewed the evidence on counselor attitudes and possible sex bias. The research findings are contradictory, but included in the limited research findings available are several studies that provide a cause for concern. The studies suggest the need for a policy related to counseling and interest inventories. In one study of vocational counseling, Thomas and Stewart (107) examined couselor attitudes toward career goals of women with what they classified as deviate (traditionally masculine) career goals versus conforming goals (education as a career, for example). The results of their study were that women who were perceived to have non-traditional or inappropriate career goals were judged in need of further counseling.

Another study by Schlossberg and Pietrofesa (96) examined the attitudes of counselors in training. They arranged interviews between counselor trainees and a coached female counselee for a counseling practicum. During the counseling session, the counselee informed the counselor she was a transfer student to the university, she was entering her junior year in college, and could not decide whether to enter the field of engineering, a "masculine" occupation, or enter the field of education, a "feminine" occupation. The interviews were tape recorded and scored for statements indicating sex bias.

Statements by the counselor were considered biased against the woman when she expressed interest in the masculine field and the counselor rejected her interest in favor of the feminine vocation. The counselor's statement was considered biased for the female counselee when she expressed interest in the masculine occupation and the counselor supported or reinforced the positive interest. Examples of the statements classified as negative bias (NB) and positive bias (PB) were:

Marriage and Family - Family attachment

(NB) Would your husband resent your being an engineer?
(PB) Being an engineer would not interfere with your becoming married.

Educational Preparation - Classes one must take to enter the field and the kinds of classes already taken

(NB) The course work in engineering is very difficult.
(PB) Your classwork up to now shows that you would do well as an engineer.

Working Conditions - Where, with whom, what kinds of work, and/or under what conditions work is done

(NB) Engineering...it is very, you know, technical, and very, I could use the term "unpeopled."

(PB) You could work at a relaxed pace as an engineer.

Masculine Occupation - Identification of occupation as masculine

(NB) You normally think of this as a man's field.
(PB) There is no such thing as a man's world anymore.

Counselors displayed more bias (used more negative bias statements) against females entering a so-called "masculine" occupation than for females entering a so-called "feminine" occupation. Also, feminine counselors displayed as much bias against females as their male counterparts. These studies (although contradictory studies exist in the research literature) are well-supported by informal statements from counselors, parents, and students. While sex bias in counselor attitudes may not be extensively documented in the research literature, there is sufficient concern that counselors, teachers, and administrators should be aware of its possibility and local policy should ensure that counseling is sex-fair.

Policy

The policies suggested here are that local schools provide a career guidance staff checklist to insure regularized procedures for counseling students and countering sex bias in existing tests, interpretative materials, and counselor behavior. Students should be given a copy of the checklist also. Categories for such a checklist have been suggested by Schiffer (95). Schiffer has grouped the major categories for such a checklist as follows: limits of tests; full range of career choices; test results; and counseling back-up for students seeking to try non-traditional jobs. These categories are briefly described below.

• Limits of the test

Schiffer suggests developing a statement on sex-bias in the career interest instrument. Problems of sex-bias that are present in the existing inventories should be explained to the student. An explanation will help to assure bias-free counseling by encouraging students to ask questions about alternatives. Secondly, there should be a statement on sex stereotyping of occupations in our culture. The vocational interests and choices of men and women are influenced by many environmental and cultrual factors, including early socialization, traditional sex-role expectations of society, conflicts of home and family versus career, and the experiences of typical women and men as members of various ethnic and social class groups. Discussion of how these factors may influence career choice and interests should assist students to understand limits in the test scores and to more readily evaluate a large range of occupational choices. Counselors should also make sure that students understand the distinction between measures of interests and ability, and the difference between interest and current knowledge about a profession. Such counseling will help to assure that students who have not obtained specific knowledge or who have had limited personal experience because of their sex will be able to pursue or consider more non-traditional interests.

• Providing a full range of career choices

Particularly for occupational scales, or other scales developed with separate sex norms, students should be given their results on both female and male norms or scales. This policy is in agreement with the NIE Guidelines, with Title IX interpretations, and is followed by most test scoring services. In presenting norms or scales for both sexes, counselors have the opportunity again to discuss the influences of past experience and sex-role socialization as they may affect student results on an interest inventory. Counselors should examine clusters of career choices, since the full range of occupations for men and women are not available.

Cole (13) has made a policy recommendation in this area. Based on her study of several interest inventories, and the limited or different occupations scaled for men and women, Cole recommends that interest inventory scales be used only to locate a woman's interests on the circular structure of a domain of occupations. Lists of both "men's" and "women's" occupations that relate to that location should then be used. The implications are that interpretative materials relating interest inventories to a broad domain of occupations are an addition to career inventories that may lessen the effect of "biased" item pools and separate criterion groups for occupations.

Another technique to be used is to discuss a student's score in terms of a ranking of interests rather than on the basis of the absolute value or magnitude of the score.

- Attention to test results

Counselors should stress student evaluation of career interest inventories in light of the student's own sense about career and activity interests. If the test does not confirm a student's interest or seems more narrow than those interests, futher exploration of careers or job possibilities are necessary. Counselors should also encourage students to seek other experiences. For students who score high on sex-traditional occupations or a narrow range of occupations, counselors may need to examine whether students have had any unusual experiences or should seek other experiences to suggest alternative careers that the student may wish to explore.

- Counseling back-up for students seeking to try non-traditional jobs

This is a critical category, since students who express interest in non-traditional jobs, whether they are female or male, may experience pressure to conform to more traditional jobs (for their sex) from peers, parents, teachers, or others with whom they may discuss their career plans. Counselors can assist and support students with non-traditional interests by stressing the current emphasis on affirmative action obligations for hiring and promotion of woemn and men in non-traditional jobs. Non-discrimination also applies in selection at institutions of higher education and financial support, areas in which there traditionally have been barriers for women's access to higher education.

If a school develops a checklist for transmitting the information described above to students, and counselors expand upon the checklist in talking to individual students, the school will have actively met Title IX requirements and will have taken a major step toward sex-fair counseling.

Another set of policy recommendations, which elaborate the suggestions above, have been suggested by Cook (cited in Schiffer, 95). These policy statements are suggested for inclusion in a checklist for counselors to monitor their own behavior. The checklists may contain policies such as the following:

(1) Counselors are equally available to male and female students on request.

(2) Male and female students or potential students are referred to counselors in approximately equal numbers.

(3) Counselors recommend programs and courses without regard to the sex of the inquiring students.

(4) Career information materials have been excluded from counseling and guidance programs when they contain sex bias and sex-role stereotyping.

(5) Career couseling programs provide role models of men and women in a variety of jobs and occupations (including those non-traditional to their sex)

(6) Men and women are equally represented on the counseling staff.

(7) Interest inventories and other appraisal instruments that contain sex bias have been eliminated from use or steps have been taken to reduce the ill effects of their bias on occupational aspiration and occupational choice.

(8) As great an emphasis is placed on the career choices and career decisions of women as on the same decisions of men.

(9) Women and men students are provided information about their rights to

equal educational and employment opportunities under the law and are provided simulated activities for dealing with sexism and discrimination.

(10) Programs are planned and conducted for parents that assist them in working with their sons and daughters on career decisions, especially with respect to sex bias and sex-role stereotyping they may encounter.

(11) Guidance and counseling and placement and follow-up records are maintained and are reviewed periodically for the differential impacts of the instructional, counseling and guidance, and placement programs on females and males who leave or complete school.

Another major policy to be considered is whether guidance counseling staff are provided with the in-service training necessary for sex-fair counseling and guidance for students. The Sex Equality in Guidance Opportunities (SEGO) materials are available for in-service training (APGA,4). Thus, there are a series of policies that can result in sex-fair counseling in guidance for students at all levels of the educational system. These include policies on the development and use of lists about the sex-fairness of interest inventories that are available for both counselors and students, developing checklists for counseling services more broadly viewed, and developing or insuring that staff participate in in-service workshops designed to assist them to provide sex-fair counseling and use of interest inventories. A related type of educational test, the aptitude test, has many of the same issues and policies as described here for the interest measures.

APTITUDE MEASURES AND SEX BIAS

OVERVIEW

The importance of examining issues of sex bias in aptitude tests cannot be over-emphasized. Aptitude tests are used in high schools and in colleges to assist in counseling students and selecting students for college entry. In addition, aptitude tests are often used in the employment selection process. One of the best known aptitude measures is the Differential Aptitude Tests (DAT) (Bennett, Seashore, and Wesman, 5). The titles of the tests in the DAT suggest the differences between educational achievement tests, interest measures, intelligence measures, and aptitude tests. The DAT has the following tests: verbal reasoning, numerical ability, abstract reasoning, clerical speed and accuracy, mechanical reasoning, space relations, spelling, and language usage. Thus, there is a wider range of tests of abilities in an aptitude battery than in an intelligence test, and achievement tests reflect school curricula more directly.

The DAT scores are reported to students on each individual test and the interpretation that is given to students is:

> When you took the Differential Aptitude Tests you were told that their purpose was to help you understand your abilities better--see where your strengths and weaknesses lie--plan your studies and think about your future career. This report will tell you how you scored and will help you use this information as you face the need to make many kinds of decisions:
>
> What courses should I elect next year? What the year after? College or not? Business or technical course? More science and math? How about languages? What career should I consider? How can I get ready for the careers that seem reasonable? Do my abilities jibe with my interests? With my opportunities? (Psychological Corporation, 89, p. 1)

The definition of aptitudes presented to counselors and students emphasizes that, "...Simply--aptitude is the capacity to learn. You take aptitude tests in order to be able to make better predictions of how you can expect to develop in school and in a job." (Psychological Corporation, 89, p. 1) The emphasis on a measure of ability as telling the individual's capacity for future learning and prediction of future attainments is what makes the scores important to students as well as those working with students.

The use of aptitude measures in high school decisions for course placement and vocational counseling, and post-secondary admissions selection and vocational counseling, as well as use in employee selection, is based upon the idea of capacity to learn and the prediction of future success. This context for interpreting aptitude scores is often generalized to the idea that an aptitude is "hereditary." Althoug the DAT Manual, (Bennett, Seashore, and Wesman, 6) provides one paragraph on the distinction between aptitude as heredity and aptitude as the result of the interaction of heredity and environment, many students, teachers and counselors probably do not maintain this distinction. Again, as with the interest measure, the emphasis on prediction is less important in a time of social change. Increasing emphasis should be given to assisting both boys and girls to explore their skills and interests related to the domains of occupations and career choices. The issues and policy recommendations below examine overt sex bias in aptitude tests, the same-sex versus combined-sex norms issue, the issue of construct validity arising from sex differences in performance, and briefly, the

issue of sex bias in selection for admission or employment.

ISSUE: OVERT SEX BIAS

Two aptitude test batteries were reviewed for overt sex bias in language usage and sex-role stereotyped portrayal of females and males. The two tests are the Differential Aptitude Tests (DAT) and the Flanagan Aptitude Classification Tests (FACT). The review of the DAT test booklets shows tht most of the tests are relatively "language free," consisting primarily of numbers and figures. Exceptions are the tests of mechanical reasoning and language usage. The test of mechanical reasoning is explained to students as:

> How easily do you grasp the common principles of physics as you see them in everyday things about you? How well do you understand the laws governing simple appliances, machinery, tools, and motions?
>
> Students who do well on the mechanical reasoning test usually like to find out how things work. They often are better than average at learning how to construct, operate, or repair complicated equipemnt...People who do poorly on this test may find the work rather hard or uninteresting in physical sciences and in those shop courses which demand thinking and planning, rather than just skill and using one's hands...Boys score considerably higher than girls on the MR and SR tests. Therefore a girl who does quite well on these tests, as compared with the average girl, may still be far below the average boy. A girl interested in mechanical or engineering work should ask her counselor to figure her MR and SR percentiles in comparison with boys as well as girls. (Psychological Corporations, 89, p.4)

The MR test has illustrations for each test item and the illustrations make very clear that this is a "man's world." The items (Form S) show jockeys, a man and a pulley, astronauts, men with mirrors, a man throwing a ball, a man with a horse and wagon, two men with hammers, pulleys and men, a man and a pulley, and two men with hoses. The issue of interpretation and meaning of the mechanical reasoning test will be discussed below. The only woman portrayed in the entire MR test is a woman in a wheelchair.

The DAT Manual (Bennett, Seashore, and Wesman, 6) very consistently refers to counselors, students, and pupils as "he." Secretaries are referred to as "she," and a majority of the examples use males to illustrate use of the DAT scores. (This is not true in the casebook, Counseling from Profiles (Psychological Corporation, 1977), where more girls than boys are used to illustrate counseling with the test.)

The second aptitude test reviewed has similar examples of overt sexism in language usage and illustrations. The FACT test booklet covers show four males and three females, apparently in sex-stereotyped occupations. The women are portrayed as a stenographer, a switchboard operator, and a filing clerk. The males are apparently a scientist, a tailor, a butcher, and a lecturer. There is a context for the process being measured (e.g., Ed and Jack are working, the foreman is a he, and so on). Similar usage appears in the test of expression. In the test of ingenuity, women serve as a hostess for a children's party and to bake a cake. The men are portrayed in occupations.

The counselor's booklet for FACT (1953) provides case studies or examples that are all male, with the exception of one female whose occupation may be that of a nurse. In summary, the FACT has overt sex bias in both language usage and illustrations, as does the DAT.

Policy

The policy of test users with these aptitude tests should be the same as discussed earlier with the educational achievement tests and interest measures. Any aptitude test should be reviewed for overt sex bias. If sex bias is found and there is an alternative, the test should not be used. If a test containing overt sex bias is to be used, a policy to ensure sex fairness should be developed. One of the main uses of aptitude measures is in counseling, and counselors should refer to the set of recommendations discussed under interest measures. Materials and specific coun-

seling discussion should inform students that all occupations are open to qualified persons of either sex, and an emphasis should be placed on developing exploratory activities so that students may check whether aptitudes reflect interest and experience. Test results should never be reported to students without counseling or interpretive materials that discuss the possible effects of experience--particulary sex-linked experiences--on aptitude measures.

ISSUE: SEX DIFFERENCES IN PERFORMANCE AND THEIR INTERPRETATION

For a variety of reasons discussed earlier, sex differences appear on some tests in aptitude batteries. For the FACT, the Technical Supplement (FACT, 1954) states that girls obtain higher scores on the expression, coding, and tables' tests, and that boys obtain higher scores on mechanics, assembly, scales, and patterns. However, there are no separate sex norms, and the interpretive materials for students do not discuss any possible sex differences.

In the Differential Aptitude Test, students check off their sex on the answer sheet, and separate sex norms are presented for grades 8 through 12, fall and spring norms. There are consistent sex differences in average scores of the clerical, spelling, language usage, and mechanical reasoning test, with girls obtaining higher scores on the first three, and boys higher on the latter. Table 10 provides a summary of the differences in raw score equivalents of the 50th percentile on the boys' and and girls' norms for four DAT tests.

As shown in the Table 10, there are few sex differences in verbal reasoning, numerical reasoning, abstract reasoning, and space relations for raw score equivalents of the 50th percentile across grades 8 through 12. The differences shown for space relations are in the usual direction, with males scoring a few raw score points higher in the later high school years.

> The item selection process is not well described for the DAT: Difficulty and discrimination values were computed for each item, by grade and sex, and these statistics were used as an aid in selecting the content of the final forms of the new test. (Bennett, Seashore, and Wesman, 6, p. 10.)

Table 10

Similarity of Raw Score Equivalents of the 50th Percentile for Females and Males on Four DAT Tests

Test	Grade Level					Total No. of items in Test
	8	9	10	11	12	
Verbal Reasoning	=[a]	=	=	=	=	50
Numerical Ability	1^{b}_{G}	1_{G}	=	=	2_{B}	40
Abstract Reasoning	=	=	=	=	1	50
Space Relations	=	=	2_{B}	4_{B}	4_{B}	60

a equal sign (=) indicates raw score equivalent of 50th percentile is the same for boys and girls

b each raw score difference was for girls (G) or boys (B) having a higher raw score, thus requiring a higher raw score for the same percentile rank as on the other-sex norms.

The small differences in raw score means on the subtests in Table 10 may be the result of a deliberate effort to provide similar distributions of ability for females and males (as is done on many intelligence tests). The differences here are very small compared to the stereotype of differences in female and male performance that abounds in the psychological and popular literature for these abilities.

The DAT reports separate sex norms for every test and battery. The FACT mentions sex differences in performance but does not provide sex-differentiated norms. The issue of concern here is which sets of norms should be reported to students and how the interpretive materials for students and counselors should describe the norms.

Policy

The difficulty of reporting combined-sex (i.e., one set of norms based on both boys and girls) is illustrated by an example given in the DAT Manual:

> ...the case of a boy, tested in the fall of tenth grade, whose raw score is 48 on the Mechanical Reasoning. If boys' and girls' scores were reported in a single distribution, a raw score of 48 would place the boy at the 66th percentile. On such evidence, a counselor might be inclined to suggest that the boy consider one of the careers requiring mechanical intelligence (sic). This course of action would be misleading since the tenth grade norms based on the boys alone show that a raw score of 48 falls at the 50th percentile. The boy is then seen as not at all superior in this ability when compared with other boys at his grade level. (Bennett, Seashore, and Wesman, 6, p. 46.)

The existence of such sex differences in performance suggests that the policy being followed with interest tests is also appropriate for aptitude measures. Both sets of norms should be reported to both females and males, and the existence of the differences in the percentile ranks obtained on some tests should be used in counseling and when using published interpretive materials to explore sex-related differences in experiences and activities that might lead to test score differences. In mechanical reasoning, for example, the current interpretation for girls is that:

> ...In counseling a girl who wishes to enter a mechanical field, her raw score on Mechanical Reasoning should be compared with the boy's norms for the same grade level. To indicate real promise in this field, she would need to score very high in relation to other girls. Assuming that she was tested in Grade 10, in the fall semester, she would need to be at the 95th percentile for girls in order to equal the 75th percentile for boys. We may anticipate, however, that with the present-day trend of girls showing interest in subjects and fields previously associated mainly with boys, the difference in scores obtained by the two sexes will gradually diminish. (Bennett, Seashore, and Wesman, 6, p. 46.)

Alternatively, Cronbach (15) has offered another approach in reviewing the suitability of the Armed Services Vocational Aptitude Battery (ASVAB 77) for use with high school students in the spring of 1978. In addition to technical difficulties with the tests, Cronbach emphasized the sex-linked shop-experience factor of some of the tests and recommended that the ASVAB not be used unless another battery is used and/or there are counselors to help interpret the ASVAB results.

Although the DAT tests are not as obviously based on shop-experience as the ASVAB, the test of mechanical reasoning is unlikely to provide as valid a measure of mechanical aptitude (or the capacity to learn) for girls. There are twice as many girls as boys in the norm groups who scored at the chance level in the test (10 percent of the girls obtained scores of 23 or less on the 3-alternative 70-item test, as compared with 5 percent of the boys in grade 8; in grades 9 through 12, the percent of boys at the chance level is from 3 percent to 1 percent, and for girls the percent drops to 5 percent and stays there for all four years). Some of the situations illustrated in the test items need to be examined on an individual basis to see if the sex differences in norms can be reduced and whether items based on illustrations of situations equally familiar to boys and girls can be

developed (as has been done very successfully in the item pools for one of interest measures).

For both the DAT and the FACT, as well as other aptitude tests, publishers should make available both same-sex and opposite-sex norms to counselors and students. Interpretive materials should discuss the effects of sex-stereotyping on experiences and thier possible relationships to test scores on the aptitude battery. Additionally, publishers and counselors may desire to try to quantify the students' estimates of experiences that may relate to their test score interpretation. This approach may be most appropriate for the tests in the DAT, for example, that have sex differences in performance: the clerical, spelling, language usage, and mechanical reasoning tests. Similarly for the FACT, the sex differences appeared on the expression, coding, tables, mechanics, assembly, scales, and patterns tests. Students might rate each test item (or the overall test) for the number of activities or hobbies they have participated in that are related to the item or test (none, a few, about average, more than most girls my age), number of school courses related to the item (test), and their expectation for performance on the test compared to other boys or girls of their age (top 25 percent, next 25 percent, and so on). These ratings can be combined into an index of test-related experiences and indicate that if there were few experiences related to the test, individuals might want to interpret their scores cautiously and seek further experience to test their own skills and interests.

Until such time as the publisher provides item analysis data in the test manual and evidence of the construct validity of the mechanical reasoning test for girls, this test should not be considered sex-fair for girls. Both sex norms should be shown to girls, possible reasons for differences explored, and care given to encourage girls to try activities that may check their interests and enlarge their experiences related to the occupations suggested as related to success on this measure. And, the test should not be used to retain the unequal enrollment of boys and girls in vocational education courses (Roby, 92). Girls should be encouraged to consider "non-traditional" educational experiences as a prelude to considering "non-traditional" occupations. Sex-role related limitations of experiences and interests should not limit women's educational and occupational options.

ISSUE: ADVERSE IMPACT OF SELECTION TESTS

As mentioned earlier, issues of sex bias also arise in the use of aptitude tests in selecting students (or counseling or otherwise encouraging their enrollment) for vocational education courses and college admissions. These situations are similar to the use of tests in employment selection settings and appear subject to the same regulations, the Uniform Guidelines on Employee Selection Procedures (Federal Register, December 20, 1977). The Uniform Guidelines go farther than earlier regulations in Title IX in defining adverse impact and setting standards for determining the fairness of tests. Title IX, as mentioned earlier, provides regulations on the use of tests in college admissions such that where there is a substantial disproportion of members of one sex in any particular course or admitted group, the school or college had to provide assurance that the disproportion was not the result of discrimination in the instrument or the applications of the instrument (i.e., a fair test and fair use of the test).

In the Unifrom Guidlines adverse impact is defined as a selection rate for any racial, ethnic or sex group that is less than four-fifths (80 percent) of the rate for the group with the highest rate. If 100 boys apply for a mechanics program and 20 girls apply, then if 50 boys are accepted and no girls, these selection rates might constitute evidence of adverse impact (the selection rate for boys was 50 percent, 1 in 2, and for girls, 0, 0 in 20 selected). If at least 8 girls had been accepted, then the selection rate for girls would have been four-fifths that of the boys, and this particular definition of adverse impact would be satisfied.

Whatever the basis for selection in the example above, whether a test, informal interview, school record, or other measure, the assessment procedure must be examined for sex bias and sex-fair use in order to

justify continued use of the procedures. Earlier discussions have recommended that procedures to establish a sex-fair test be defined, and at minimum they include examining the test for overt sex bias in terms of language usage, sterotyped portrayals of females and males, and positive representation of women; for item bias by providing empirical data based on one of the item analysis procedures proposed to date; and, where sex differences in average performance remain, for test bias by providing further evidence of the construct validity of the test. This series of reviews, judgments, and examination of empirical data will permit a test user to decide it the test is fair to a group of women in a particular application of the test and then examine the use of the test for sex fairness.

Fair test use has been the subject of considerable theoretical work in a limited, psychometric framework (see the Vol. 13, No. 1, 1976 issue of Journal of Educational Measurement devoted to test bias; also Linn, 64, and Novick and Ellis, 79). The psychometric approach has been to examine the predictive validity of the test for different groups. Some of the work has been done with college admissions tests such as the Scholastic Aptitude Test (Cleary, 10, for example) and other work has been in the employment selection setting (Einhorn and Bass, 27). However, in both selection settings the focus has been to examine the relationship between the test and a criterion of success--such as first-year college grades or successful performance on a job. Wild and Dwyer (117) have recommended that both the criterion and the predictor (test) be examined for their relationship to educational goals and intermediate criteria such as grades (or in employement, satsifactory performance on the job), as well as for their reliability for both females and males, and for sex bias.

The examination of test bias in selection using predictive validity has compared the validity coefficients of various proportions of groups selected under different assumptions. There appear to be two limitations to this conception of test bias. From one perspective, the test bias models so far have relied on group categories, such as racial/ethnic groups or sex, and examined whether the test functioned in a fair manner for all the individuals in a group. From another perspective, the definition of test bias has relied heavily on predictive validity to examine test fairness and not placed enough emphasis on content and construct validity.

Criticisms of the test bias models using group analyses have suggested that we are typically concerned with providing equal opportunity to individuals, for education or employment (Novick and Ellis, 79). In providing special admissions for minority groups and women to law or medical schools, we are trying to compensate for earlier disadvantage based on discrimination because of race or sex. The problem, according to Novick and Ellis, is that the disadvantage based on discrimination does not occur equally to all individuals within the race or sex group. The purpose of the compensatory policy may not be met unless there is some way of attaching an index of "educational or equal opportunity disadvantage" based on race, sex, social class, and so on, to each individual considered for hiring or admissions. This argument adds to the predictive model an explicit utility or value to hiring or enrolling individuals who are disadvantaged based on prior discrimination. Test scores are still used, with probabilities or estimates of successful performance in school, but are adjusted by the values attached to providing an education or job to individuals who have been discriminated against and thus denied equal opportunity to the point of the decision being made. This type of procedure has been formally used before in providing bonus points to veterans in civil service employment (points are added just for being a veteran, on top of whatever test scores or credential evaluations are made). Similarly, the system has operated informally in college admissions where preference is given to admitting sons and daughters of alumni and to those who may provide substantial funds to a college or to those who are from different geographic regions of the country or who have special skills (artists or athletes, for example).

The use of race or ethnicity as a selection variable to a professional school was examined by the U.S. Supreme Court in the Bakke case for medical school admission. That decision apparently accepts the use of race or other indices of disadvantage so long as there is no specific number of admissions allocated to one group or an-

other. Thus there will be further efforts to change the narrow psychometric prediction model to a model that includes both probabilities of success (prediction) and utilities. The prediction problem does not affect women as much as minorities. The "adverse impact" criterion of sex bias makes clear that the problem for women is encouraging them to enter the applicant pool.

However, from the sex bias perspective and examining educational tests, there is most cause for concern with tests of mathematics, and spatial areas, since sex differences in average performance may be found in these tests. As mentioned earlier, there may be some merit to trying to quantify an index of "mechanical experiences" (disadvantage based on sex discrimination in the Novick and Ellis sense) or "mathematics experiences" or "spatial experiences" to assist in interpreting test scores for individual women and for use on an experimental basis in selecting women for schools or fields of study that are traditionally male dominated, e.g., auto mechanics, engineering, architecture, physics, and so on. This approach would be consonant with the second criticism of the work in test bias in selection to date, the criticism that predictive validity has been too much the focus and the concept of validity needs to be broader and emphasize construct validity (Manning, 69; Messick, 75).

Manning (69) has examined recent court decisions related to test bias for the type of validity permitted in the legal judgments. Although professional test standards have emphasized predictive validity, the courts have adopted a legal standard that evidence of content or construct validity is also appropriate. And Manning suggests that the logical, hypothetical, and deductive processes of scientific inference needed for content and construct validity provide a necessary balance to predictive validity. The Uniform Guidelines provide the same view, and give definitions of both types of validity: content validity is demonstrated by data showing that a selection procedure is a representative sample of important work behaviors to be performed on the job; construct validity is demonstrated by data showing that the selection procedure measures the degree to which candidates have identifiable characteristics that have been determined to be important for successful job performance.

Thus, the Uniform Guidelines suggest that where evidence of adverse impact of test use exists, schools and employers cannot rely on predictive validity coefficients as reported in the manual for tests such as the DAT. In selecting or recommending students for advanced science courses, for vocational courses, for professional schools of law and engineering, based on tests, the test user will need to examine the sex fairness of the test, including its content and construct validity, and the sex fairness of its uses in selection or counseling. There are policy implications of this issue.

Policy

The issue of sex-fair aptitude tests and fair use of these tests has implications for school and employer policy in establishing review procedures for tests and for maintaining test data. The review procedures for aptitude tests are similar to those for achievement and interest measures and consist of forming groups to review the test items and accompanying manuals and student interpretive materials from the standpoint of sex fairness for the particular group of women who will take the tests. Judgments based on the review of test materials and test data, either collected by the publisher or locally, are part of the set of procedures to determine sex fairness. Where sex differences remain, and where the test is used in selection, further analysis of the job or performance related characteristics of the test (content and construct validity) is required. Part III of the Uniform Guidelines (1977) provides a detailed listing of all the descriptive information and data to be given for each type of validity.

The test user should maintain record that provide information on the applicant groups, and those finally selected, by sex and racial/ethnic groups designated as black, American Indian, Asian, Hispanic, and white (Caucasian other than Hispanic). These classifications are also those used for developing and monitoring affirmative action programs and could also be used for the new VEA requirements to monitor and affirm equal access to vocational educational programs.

Other policies, such as those described with the use of interest measures and sex-fair counseling, will require additional data and records and may be appropriate for aptitude measures also.

CONCLUDING REMARKS

This paper on equity for women in educational testing has focused on issues of sex bias in educational achievement tests, career interest inventories, and aptitude tests. Where each issue was raised and analyzed for implications of sex bias, policies have been suggested that are within the authority and responsibility of local and state educational administrators, teachers, counselors, parents, and even students.

One outcome of the review is the conclusion that sex bias or sex-fairness does not exist as a general quality of any educational test, any more than test reliability or validity exist as a general quality of the test. Hence there is a need for an understanding at the policy-making level in education that there will need to be a series of procedures, checklists, and operational policies that will consider sex bias and sex-fairness in state and local use of educational tests. Policy is needed to establish the use of the appropriate set of procedures in each testing situation.

This view recognizes that there is no one estimate or figure that will say that a test is sex-fair. Rather, there are in the process of development what will become commonly accepted codes or sets of rules that, when applied to the evaluation of the sex-fairness of a given educational test in a particular context, will permit the determination whether the test and its use will be sex-fair. Evidence for the codification of the set of rules are in the proposed common regulations of the Civil Service Commission, the Equal Opportunity Commission, the Department of Justice, and the Department of Labor--the Uniform Guidelines on Employee Selection Procedures (Federal Register, December 20, 1977) and in the NIE Guidelines for the Assessment of Sex Bias and Sex-Fairness in Career-Interest Inventories. These Guidelines are more explicit with respect to test bias and sex-fairness than the current professional standards as codified in the Standards for Educational and Psychological Tests (American Psychological Association, 101). It is likely that the Standards will be brought more in line with federal regulations and the NIE Guidelines when they are revised in the near future.

Distinctions have been made between the set of procedures (reviews, data analyses, and judgments) that will assist in the evaluation of whether a test is sex-fair for a particular group of women or girls in a particular setting, and whether there is sex-fair use of the test. In the first instance the test will be reviewed by the publisher in the test development process and by local committees for overt sex bias in language usage, stereotyped portrayal of men and women, and positive representation of women in all areas of culture. Similar reviews will examine the test administration manual, technical manual, and all interpretive materials provided to test administrators, counselors, teachers, and students. Other evidence of sex-fairness of the test will be in the empirical data presented, including the statistical analyses of items for bias, decision rules for selecting items, and the average performance of females and males on the test. Where sex differences in performance are found, evidence of the construct validity of the test for women is required. Issues concerning the types of norms that are required when mean sex differences occur were discussed and policy recommendations made.

In sex-fair test use in counseling, the use of achievement test scores, career-interest inventories, and aptitude tests were discussed. Of particular concern are course enrollment, vocational preparation, college major, and labor force participation data showing girls and women in sex-segregated areas--under-represented in crafts and trades, mathematics, and sciences, and in major professions such as business, law, and medicine. Procedures and policies are recommended to encourage women to enter non-traditional areas and to ensure that educational testing does not contribute to inequity in educational opportunities.

The sex-fair use of aptitude tests in course, program, and employment selection was also examined. A changing emphasis from predictive validity to requiring job- and performance-related analyses of tests and criteria is seen as a positive move away from a narrow perspective of test bias. Logical judgments and careful development of the rationale for any use of aptitude tests where adverse impact is found should permit a better analysis of the role of past experience and sex-role expectations and stereotypes for any possible effect on the observed performance of girls and women on educational tests.

While the emphasis in the review has been on issues and policy in relation to sex bias and educational tests, the sets of procedures suggested for local review and data analysis will contribute to the better use of educational tests in general. Too often tests are used and scores obtained and interpreted without a consciousness of the individual items in the test and an examination of student performance at the item level. This focus, it is argued here, will contribute to the integration of tests with curriculum and instruction on a broader basis. In this sense, considerations of educational equity for women contribute to improved education, a worthy goal for all concerned with the educational process.

REFERENCES+

1. Almquist, E.M. Women in the Labor Force: Review Essay. Signs, 1977, Vol. 2, Pp. 843-855.

2. Angoff, W.H. The Investigation of Test Bias in the Absence of an Outside Criterion. Paper presented at the National Institute of Education Conference on Test Bias, Washington, D.C., December 4, 1975.

3. Angoff, W.H., and Ford. S.F. Item-Race Interaction on a Test of Scholastic Aptitude. Research Bulletin 71-59. Princeton, N.J.: Educational Testing Service, October 1971.

4. APGA, The Project on Sex Equality in Guidance Opportunities (Revised Edition). Washington, D.C.: American Personnel and Guidance Association (n.d.).

5. Bennett, G.K., Seashore, H.G., and Wesman, A.G. Differential Aptitude Tests, Form S. New York: Psychological Corp., 1972.

6. Bennett, G.K., Seashore, H.G., and Wesman, A.G. Fifth Edition Manual for the Differential Aptitude Tests, Forms S and T. New York: Psychological Corp., 1974.

7. Brim, O.G., Jr., Glass, D.O., Neulinger, J., and Firestone, I.J. American Beliefs and Attitudes about Intelligence. New York: Russell Sage Foundation, 1969.

8. Campbell, J. Differential Response for Female and Male Law Students on the Strong-Campbell Interest Inventory: The Question of Separate Sex Norms. Journal of Counseling Psychology, 1976, Vol. 23, Pp. 130-135.

9. Cardall, C., and Coffman, W.E. A Method for Comparing the Performance of Different Groups on the Items in a Test. Research Bulletin 64-61. Princeton, N.J.: Educational Testing Service, 1964.

10. Cleary, T.A. Test Bias: Prediction of Grades for Negro and White Students in Integrated Colleges. Journal of Educational Measurement, 1968, Vol. 5, Pp. 115-124.

11. Coffman, W.E. Sex Differences in Response to Items in an Aptitude Test. Eighteenth Yearbook, National Council on Measurement in Education, 1961, Pp. 117-124.

12. Cole, N.S. Approaches to Examining Bias in Achievement Test Items. Paper presented at the annual meeting of the American Personnel and Guidance Association, Washington, D.C., March 1978.

13. Cole, N.S. On Measuring the Vocational Interests of Women. Journal of Counseling Psychology, 1973, Vol. 20, Pp. 105-112.

14. Cooper, J.F. Comparative Impact of the SCII and the Vocational Card Sort on Career Salience and Career Exploration of Women. Journal of Counseling Psychology, 1976, Vol. 23, Pp. 348-352.

15. Cronbach. L.J. (cited in) Caution Urged in Use of Armed Forces Battery. Guidepost, February 16, 1978, Vol. 20, No. 12, Pp. 1, 10.

+Items followed by an ED number (for example, ED 072 368) are available from the ERIC Document Reproduction Service (EDRS). Consult the most recent issue of Resources in Education for the address and ordering information.

16. Deckard, B. The Women's Movement: Political, Socioeconomic, and Psychological Issues. New York: Harper and Row, 1975.

17. Denker, E.R., and Tittle, C.K. "Reasonableness" of KOIS for Re-entry Women: Implications for Test Validity. Educational and Psychological Measurement, 1976, Vol. 36, Pp. 495-500.

18. Dewey, C.R. Exploring Interests: A Non-Sexist Method. Personnel and Guidance Journal, 1974, Vol. 52, Pp. 311-315.

19. Diamond, E.E. (Ed.) Issues of Sex Bias and Sex Fairness in Career Interest Measurement. Washington, D.C.: U.S. Department of Health, Education and Welfare (available from the National Institute of Education, Washington, D.C., 20208), 1975.

20. Diamond, E.E. Minimizing Sex Bias in Achievement Testing. Measurement and Evaluation in Guidance, 1975, Pp. 28-34.

21. Donlon, T.F. Content Factors in Sex Differences on Test Questions. Paper presented at the meeting of the New England Educational Research Organization, Boston, June 1971.

22. Donlon, T.F., Ekstrom, R.B. Lockheed, M., and Harris, A. Final Report. Performance Consequences of Sex Bias in the Content of Major Achievement Batteries. PR77-11. Princeton, N.J.: Educational Testing Service, July 1977.

23. Draba, R.E. The Identification and Interpretation of Item Bias. Education Statistics Laboratory Memorandum 26, March 1977. Chicago: University of Chicago, 1977.

24. Durovic, J.J. Definitions of Test Bias: A Taxonomy and an Illustration of an Alternate Model. Dissertation Abstracts International, 1975, Vol. 36, No. 5.

25. Durovic, J.J. Use of the Rasch Model in Assessing Item Bias. Paper presented at the annual meeting of the Eastern Educational Research Association, Williamsburg, Va., March 1978.

26. Echternacht, G. A Quick Method for Determining Item Bias. Educational and Psychological Measurement, 1974, Vol. 34, Pp. 271-280.

27. Einhorn, H.J., and Bass, A.R. Methodological Considerations Relevant to Discrimination in Employment Testing. Psychological Bulletin, 1971, Vol. 75, Pp. 261-269.

28. Ekstrom, R.B. Barriers to Women's Participation in Post-Secondary Education: A Review of the Literature. Research Bulletin 72-49. Princeton, N.J.: Educational Testing Service, 1972. ED 072 368.

29. FACT Counselor's Booklet. Chicago: Science Research Associates, 1953.

30. FACT Technical Supplement. Chicago: Science Research Associates, 1954.

31. Fishbein, R.L. An Investigation of the Fairness of the Items in a Test Battery. Paper presented at the annual meeting of the National Council on Measurement in Education, Washington, D.C., April 1975.

32. Fishbein, R.L. The Fairness of Test Items for Statistically Equated Groups. Paper presented at the annual meeting of the American Educational Research Association, New York, April 1977.

33. Flaugher, Ronald L. The Many Definitions of Test Bias. American Psychologist. Vol. 33, No. 7, July 1978. Pp. 671-679.

34. Frasher, R., and Walker, A. Sex Roles in Early Reading Textbooks. The Reading Teacher, 1972. Vol. 25, Pp. 741-749.

35. Goslin, D.A. Teachers and Testing. New York: Russell Sage Foundation, 1967.

36. Gottfredson, G.D., and Holland, J.L. Vocational Choices of Men and Women: A Comparison of Predictors from the Self-Directed Search. Journal of Counseling Psychology, 1975. Vol. 22, Pp. 28-34.

37. Grambs, J.P. Sex-Stereotypes in Instructional Materials, Literature, and Language: A Survey of Research.

Women Studies Abstracts, 1972, Vol. 1, Pp. 1-4, 91-94.

38. Green, D.R. Racial and Ethnic Bias in Test Construction. Monterey, Calif.: CTB/McGraw Hill, 1972.

39. Green, D.R., Draper, J.F. Exploratory Studies of Bias in Achievement Tests. Paper presented at the annual meeting of the American Psychological Association, Honolulu, September 1972.

40. Gump, J.P., and Rivers, L.W. A Consideration of Race in Efforts to End Sex Bias. In E.E. Diamond (Ed.), Issues of Sex Bias and Sex Fairness in Career-Interest Measurement. Washington, D.C.: U.S. Department of Health, Education and Welfare (available from the National Institute of Education, Washington, D.C. 20208), 1975.

41. Hambleton, R.K., and Cook, L.L. Latent Trait Models and their Use in the Analysis of Educational Test Data. Journal of Educational Measurement, 1977, Vol. 14, Pp. 75-96.

42. Hansen, J.C. Exploring New Directions for Strong-Campbell Interest Inventory Scale Construction. Journal of Vocational Behavior, 1976, Vol. 9, Pp. 147-160.

43. Hansen, J.C. Sex Differences in Vocational Interests: Three Levels of Exploration. In C.K. Tittle and D.G. Zytowski (Ed.), Sex-Fair Interest Measurement: Research and Implications. Washington, D.C.: U.S. Department of Health, Education and Welfare (available from the National Institute of Education, Washington, D.C. 20208), 1978.

44. Hanson, G.R., and Rayman, J. Validity of Sex-Balanced Interest Inventory Scales. Journal of Vocational Behavior, 1978, in press.

45. Hanson, G.R., Noeth, R.J., and Prediger, D.J. The Validity of Diverse Procedures for Reporting Interest Scores: An Analysis of Longitudinal Data. Journal of Counseling Psychology, 1977, in press.

46. Harmon, L.W., and Conroe, F.L. Sex Stereotyping in Interest Items. Measurement and Evaluation in Guidance, 1976, Vol. 8, Pp. 215-220.

47. Harway, M., and Astin, H.S. Sex Discrimination in Career Counseling and Education. New York: Praeger, 1977.

48. Holland, J.L. The Use and Evaluation of Interest Inventories and Simulations. In E.E. Diamond (Ed.), Issues of Sex Bias and Sex Fairness in Career Interest Measurement. Washington, D.C.: U.S. Department of Health, Education, and Welfare (available from the National Institute of Education, Washington, D.C. 20208). 1975.

49. Holland, J.L. Toward Definable and Beneficial Resolutions of the Interest Inventory Controversy. Baltimore, Md.: Department of Social Relations, Johns Hopkins University, 1977.

50. Holland, J.L., Takai, R., Gottfredson, G.D., and Hanau, C. A Multivariate Analysis of the Effects of the Self-Directed Search on High School Girls. Journal of Counseling Psychology, 1978, in press.

51. Holmen, M.G., and Docter, R.F. Educational and Psychological Testing: A Study of the Industry and its Practices. New York: Russell Sage Foundation, 1972.

52. Ironson, G.H. A Comparative Analysis of Several Methods of Assessing Item Bias. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.

53. Jacobs, C., and Eaton, C. Sexism in the Elementary School. Today's Education, 1972, Vol. 61, No. 9, Pp. 20-22.

54. Jensen, M., and Beck, M.D. Gender Balance Analysis of the Metropolitan Achievement Tests, 1978 Edition. Paper presented at the annual meeting of the National Council on Measurement in Education, Toronto, March 1978.

55. Johansson, C.B., and Harmon, L.W. Strong Vocational Interest Blank: One Form or Two? Journal of Counseling Psychology, 1972, Vol. 19, Pp. 404-410.

56. Johnson, R.W. Relationship between Female and Male Interest Scales for the Same Occupation. Journal of Vocational Behavior, 1977, Vol. 11, Pp. 239-252.

57. Journal of Educational Measurement. Applications of Latent Trait Models, 1977, Vol. 14, No. 2.

58. Journal of Educational Measurement. On Bias in Selection, 1976, Vol. 13, No. 1.

59. Key, M.R. The Role of Male and Female in Children's Books--Dispelling All Doubt. Wilson Library Bulletin, 1971, Vol. 46, Pp. 167-176.

60. King, P.E., and Blount, H.P. The Influence of Sex-Typed, Verbal Math Test Problems in Performance of Sixth Grade Students. Paper presented at the annual meeting of the National Council on Measurement in Education, Washington, D.C., April 1975.

61. Kirkland, M.C. The Effects of Tests on Students and Schools. Review of Educational Research, 1971, Vol. 41, Pp. 303-350.

62. Lamb, R.R. Validity of the ACT Interest Inventory for Minority Group Members. In C.K. Tittle and D.G. Zytowski (Eds.), Sex-Fair Interest Measurement: Research and Implications. Washington, D.C.: U.S. Department of Health, Education, and Welfare (available from the National Institute of Education, Washington, D.C. 20208), 1978.

63. Lee, Arthur M., and Fitzgerald, Dorris. Learning a Living Across the Nation. Volume V. Project Baseline. Fifth National Report. Baseline Year: 1974-75 (Fiscal Year 1975). Part 2: Statistical Almanac. Flagstaff, Arizona: Northern Arizona University, November 1976. ED 137 551.

64. Linn, R.L. Fair Test Use in Selection. Review of Educational Research, 1973, Vol. 43, Pp. 139-161.

65. Lord, F.M. Practical Applications of Item Characteristic Curve Theory. Journal of Educational Measurement, 1977, Vol. 14, Pp. 117-138.

66. Lunneborg, P.W. Interpreting Other-Sex Scores on the Strong-Campbell Interest Inventory. Journal of Counseling Psychology, 1975, Vol. 22, Pp. 494-499.

67. Maccoby, E.E., and Jacklin, C.N. The Psychology of Sex Differences. Stanford, Calif.: Stanford University Press, 1974.

68. Malcolm, A.H. Most Common Verb in Schools, Study Finds, Is...Is. New York: New York Times, September 4, 1971, P. 22.

69. Manning, W.H. Educational Research, Test Validity, and Court Decisions. Paper presented at the annual meeting of the American Educational Research Association, New York, April 1977.

70. Mathews, W.M. The Development of Sexism in American Education. Paper presented at the annual meeting of the American Educational Research Association, Chicago, April 1972.

71. McCarthy, K. Sex Bias in Tests of Mathematical Aptitude. Ph.D. Dissertation, City University of New York, 1975.

72. Medley, D.M., and Quirk, T.J. The Application of a Factorial Design to the Study of Cultural Bias in General Culture Items in the National Teacher Examination. Journal of Educational Measurement, 1974, Vol. 11, Pp. 235-245.

73. Merz, W.R. Test Fairness and Test Bias: A Review of Procedures. Paper presented at an Office of Education Conference on Achievement Testing of Disadvantaged and Minority Students for Educational Program Evaluation, Reston, Va., May 1976.

74. Merz, W.R., and Rudner, L.M. Bias in Testing: A Presentation of Selected Methods. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.

75. Messick, S. The Standard Problem: Meaning and Values in Measurement and Evaluation. Presidential address to the Division of Evaluation and Measurement of the American Psychological Association, New Orleans, August 1974.

76. Milton, G.A. Five Studies of the Relation between Sex-Role Identification and Achievement in Problem Solving. Technical Report 3. New Haven, Conn.: Yale University, 1958.

77. Milton, G.A. The Effects of Sex-Role Identification upon Problem-Solving Skill. Journal of Abnormal and Social Psychology, 1957, Vol. 55, Pp. 208-212.

78. Mowder, B.A., and Chase, C.I. Sex-Biased Mathematics Items and Performance of Elementary School Students. Paper presented at the annual meeting of the National Council on Measurement, New York, April 1977.

79. Novick, M.R., and Ellis, D.D., Jr. Equal Opportunity in Educational and Employment Selection. American Psychologist, 1977, Vol. 32, Pp. 306-320.

80. Now Times, Report on Working Women. October 1978, P. 12.

81. Oliver, L.W. Evaluating Career Counseling Outcome for Three Modes of Test Interpretation. Measurement and Evaluation in Guidance, 1977, Vol. 10, Pp. 153-161.

82. Perl, H. Women Still Lag in Work Income. National Now Times, July 1978, P. 11.

83. Picou, J.S., and Campbell, R.E. Career Behavior of Special Groups: Theory, Research, and Practice. Columbus, Ohio: Charles E. Merrill, 1975.

84. Plake, B.S., Hoover, H.D., and Loyd, B.H. An Investigation of Differential Item Performances by Sex on the Iowa Tests of Basic Skills. Paper presented at the annual meeting of the National Council on Measurement in Education, Toronto, March 1978.

85. Prediger, D.J., and Cole, N.S. Sex-Role Socialization and Employment Realities: Implications for Vocational Interest Measures. Journal of Vocational Behavior, 1975, Vol. 7, Pp. 239-251.

86. Prediger, D.J., and Hanson, G.R. Some Consequences of Using Raw-Score Reports of Vocational Interests. Journal of Educational Measurement, 1977, Vol. 4, Pp. 323-333.

87. Prediger, D.J., and McLure, G.T. Final Report: The Exploration of Personally Relevant Career Options in Science and Technology. Iowa City: American College Testing Program, October 1976.

88. Psychological Corporation. Counseling from Profiles: A Casebook for the Differential Aptitude Tests (Second Ed.). New York: Psychological Corp., 1977.

89. Psychological Corporation. Your Aptitudes as Measured by the Differential Aptitude Tests, Forms S and T. New York: Psychological Corp., 1973.

90. Rayman, J.R. Sex and the Single Interest Inventory: The Empirical Validation of the Sex-Balanced Interest Inventory Items. Journal of Counseling Psychology, 1976. Vol. 23, Pp. 239-246.

91. Reed, C.W. Statistical Issue Raised by Title IX Requriements on Admission Procedures. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1976.

92. Roby, P.A. Toward Full Equality: More Job Education for Women. School Review, 1976, Vol. 84, Pp. 181-211.

93. Saario, T.N., Jacklin, C.N., and Tittle, C.K. Sex-Role Stereotyping in the Public Schools. Harvard Educational Review, 1973, Vol. 43, Pp. 386-416.

94. Samuda, R.J. The Psychological Testing of American Minorities. New York: Dodd, Mead, 1975.

95. Schiffer, L. Legal Issues Regarding Sex Bias in Selection and Use of Career Interest Inventories. In C.K. Tittle and D.Z. Zytowski (Eds.), Sex-Fair Interest Measurement: Research and Implications. Washington, D.C.: U.S. Department of Health, Education and Welfare (available from the National Institute of Education, Washington, D.C. 20208), 1978.

96. Schlossberg, N.K., and Pietrofesa, J.J. Perspectives on Counselors' Bias: Implications for Counselor Education.

The Counseling Psychologist, 1973, Vol. 4, Pp. 44-54.

97. Schueneman, J. Ethnic Group Bias in Intelligence Test Items. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.

98. Schueneman, J. Further Considerations in the Assessment of Bias Interest Items. Paper presented at the annual meeting of the American Psychological Association, Toronto, August 1978.

99. Schueneman, J. Validating a Procedure for Assessing Bias in Items in the Absence of an Outside Criterion. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1976.

100. Sherman, S.W. Multiple-Choice Test Bias Uncovered by Use of an "I Don't Know" Alternative. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1976.

101. Standards for Educational and Psychological Tests. Washington, D.C.: American Psychological Association, 1974.

102. Stebbins, L.B., Ames, N.L., and Rhodes, I. Sex Fairness in Career Guidance. Cambridge, Mass.: Abt Publications, 1975.

103. Strainchamps, E. Our Sexist Language. In V. Gornick and B.K. Moran (Eds.) Woman in Sexist Society. New York: Basic Books, 1971.

104. Strassberg-Rosenberg, B., and Donlon, T.F. Content Influences on Sex Differences in Performance on Aptitude Tests. Paper presented at the annual meeting of the National Council on Measurement in Education, Washington, D.C., April 1975.

105. Super, D.E., and Hall, D.T. Career Development: Exploration and Planning. Annual Review of Psychology, 1978, Vol 29, Pp. 333-372.

106. Takai, R., and Holland, J.L. The Relative Influence of the Vocational Card Sort, the Self-Directed Search, and the Vocational Exploration and Insight Kit on High School Girls. (Submitted for publication, 1978.)

107. Thomas, A.H., and Stewart, N.R. Counselor Response to Female Clients with Deviate and Conforming Career Goals. Journal of Counseling Psychology, 1971, Vol. 18, Pp. 352-357.

108. Thorne, B., and Henley, N. (Eds.) Language and Sex: Difference and Dominance. Rowley, Mass.: Newbury House, 1975.

109. Tittle, C.K. Fairness in Educational Achievement Testing. Education and Urban Society, 1975, Vol. 8, Pp. 86-103.

110. Tittle, C.K., and Denker, E.R. KOIS Profiles of Re-entry Women. Journal of Counseling Psychology, 1977, Vol. 24, Pp. 293-300.

111. Tittle, C.K., and Zytowski, D.G. (Eds.) Sex-Fair Interest Measurement: Research and Implications. Washington, D.C.: U.S. Department of Health, Education and Welfare (available from the National Institute of Education, Washington, D.C. 20208), 1978.

112. Tittle, C.K., McCarthy, K.A., and Steckler, J.F. Women and Educational Testing: A Selective Review of the Research Literature and Testing Practices. Princeton, N.J.: Educational Testing Service and the Association for Measurement and Evaluation in Guidance, 1974.

113. U.S. Working Woman: A Chartbook (Bulletin 1880). Washington, D.C.: U.S. Department of Labor, Bureau of Labor Statistics, 1975.

114. Veale, J., and Foreman, D.I. Cultural Variation in Criterion-Referenced Tests: A "Global" Item Analysis. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, April 1976.

115. Vetter, B.M. Women and Minority Scientists. Science, 1975, Vol. 189, P. 751.

116. Webber, P., and Harmon, L.W. The Reliability and Concurrent Validity of Three Types of Occupational Scales for Two Occupational Groups: Some Evidence Bearing on Handling Sex

Difference in Interest Scale Construction. In C.K. Tittle and D.G. Zytowski (Eds.), Sex-Fair Interest Measurement: Research and Implications. Washington, D.C.: U.S. Department of Health, Education and Welfare (available from the National Institute of Education, Washington, D.C. 20208), 1978.

117. Wild, C.L., and Dwyer, C.A. Sex Bias in Selection. RM 78-1. Princeton, N.J.: Educational Testing Service, February 1978.

118. Wright, B.D., Mead, R., and Draba, R. Detecting and Correcting Test Bias with a Logistic Response Model. Research Memorandum No. 22. Statistical Laboratory, Department of Education, University of Chicago, October 1976.

119. Zytowski, D.G. The Effects of Being Interest-Inventoried. Journal of Vocational Behavior, 1977, Vol. 11, Pp. 153-157.